

Accepted for publication in The Astrophysical Journal

HI CONTENT AND OPTICAL PROPERTIES OF FIELD GALAXIES FROM THE ALFALFA SURVEY. I. SELECTION OF A CONTROL SAMPLE

M. Carmen Toribio and José M. Solanes

Departament d'Astronomia i Meteorologia and Institut de Ciències del Cosmos, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain

`mctoribio@am.ub.es, jm.solanes@ub.edu`

and

Riccardo Giovanelli and Martha P. Haynes

Center for Radiophysics and Space Research, Space Sciences Building, Cornell University, Ithaca, NY 14853

National Astronomy and Ionosphere Center, Cornell University, Ithaca, NY 14853

`riccardo@astro.cornell.edu, haynes@astro.cornell.edu`

and

Karen L. Masters

Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Burnaby Road, Portsmouth, P01 3FX, UK

`karen.masters@port.ac.uk`

ABSTRACT

We report results from a study of the HI content and stellar properties of nearby galaxies detected by the Arecibo Legacy Fast ALFA blind 21-cm line survey and the Sloan Digital Sky Survey. We consider two declination strips covering a total area of $9 \text{ hr} \times 16 \text{ deg}$ in the general direction of the Virgo Cluster. The present analysis focuses on gas-rich galaxies expected to show little or no evidence of interaction with their surroundings. We seek to assemble a

control sample suitable for providing absolute measures of the HI content of gaseous objects, as well as to study the relationship between HI emission and widely-used optical measures of morphology. From a database which includes more than 15,000 HI detections, we have assembled three samples that could provide adequate HI standards. The most reliable results are obtained with a sample of 5647 sources found in low density environments, as defined by a nearest neighbor approach. The other two samples contain several hundred relatively isolated galaxies each, as determined from standard isolation algorithms based either on a combination of spectroscopic and photometric information or solely on photometric data. We find that isolated objects are not particularly gas-rich compared to their low-density-environment counterparts, while they suffer from selection bias and span a smaller dynamic range. All this makes them less suitable for defining a reference for HI content. We have explored the optical morphology of gaseous galaxies in quiet environments finding that, within the volume surveyed, the vast majority of them display unequivocal late-type galaxy features. In contrast, bona fide gas-rich early-type systems account only for a negligible fraction of the 21-cm detections. We argue that HI emission provides the most reliable way to determine the morphological population to which a galaxy belongs. We have also observed that the color distribution of flux-limited samples of optically-selected field HI emitters does not vary significantly with increasing distance, while that of non-detections becomes notably redder. This result suggests that the colors and HI masses of gas-rich galaxies cannot be very closely related.

Subject headings: PACS: 98.52.Cf, 98.52.Nr, 98.62.Lv, 98.62.Py, 98.62.Qz, 98.62.Ve

1. INTRODUCTION

One of the defining characteristics of recent extragalactic research is the execution of wide-area galaxy surveys, which have in many ways revolutionized our understanding of both large-scale structure and galaxy evolution. Most of them have been conducted in the optical/near-infrared regime, where stellar light dominates. Multicolor imaging surveys of virtually the entire sky, some of which contain over a million extended sources, notably the Two Micron All Sky Survey (2MASS; Skrutskie et al. 2006), have been complemented by spectroscopic surveys, such as the fifth edition of the Sloan Digital Sky Survey (SDSS) Quasar Catalog (Schneider et al. 2010), based upon the SDSS Seventh Data Release (DR7), which ranges in redshift from the local universe to near the reionization epoch ($z \lesssim 6$) when

the first galaxies may have appeared.

The improvement in technological capabilities over the past decade has enabled the astronomical community to build large homogeneous samples of extragalactic objects in other regions of the electromagnetic spectrum as well—the radio wavelength regime being one of the most prolific in this respect—, allowing us to develop a more thorough account of the extragalactic realm. While early blind neutral hydrogen (HI) surveys, such as the Arecibo HI Strip Survey (Zwaan et al. 1997) or the Arecibo Dual-Beam Survey (Rosenberg & Schneider 2002), yielded useful estimates of the HI mass function from shallow samples containing a relatively modest number of detections, the advent of focal plane arrays at 21-cm enabled to carry out wide solid angle surveys that led to a significant increase in the depth and, especially, size of the databases. Between 1997 and 2001, the HI Parkes All-Sky Survey (HIPASS; Meyer et al. 2004) used a thirteen-beam array to sample the entire southern sky and the northern up to Decl. $< +25^\circ$, with a beam of 15.5 arcmin and a sensitivity of 17 mJy per beam at a spectral resolution of 18 km s $^{-1}$. With a velocity range of -1280 to $12,700$ km s $^{-1}$ and more than 4000 HI sources identified in the southern sky alone, HIPASS was the first large untargeted HI survey that approached the dimensions necessary for a fair sampling of the local volume.

Currently, a seven-beam array named ALFA (Arecibo L-band Feed Array), installed at the Arecibo telescope in 2004, is being used to carry out the Arecibo Legacy Fast ALFA Survey (ALFALFA; Giovanelli et al. 2005). Its aim is to map, in the HI line, 7000 square degrees of the high galactic latitude sky between 0° and $+36^\circ$ in declination. ALFALFA offers significantly higher angular (FWHM $\sim 3'.5$) and spectral (~ 5.5 km s $^{-1}$) resolution, as well as higher sensitivity ($\sigma_{\text{rms}} = 2.2$ mJy/beam at 10 km s $^{-1}$ resolution) than any previous survey of its kind. It is also deeper, with a radial extent ($-2000 \lesssim cz_{\text{hel}} \lesssim 18,000$ km s $^{-1}$) and a median depth ($cz_{\text{hel}} \simeq 8000$ km s $^{-1}$) that should be sufficient to sample satisfactorily the largest structures seen in numerical simulations in a Λ CDM world model ($\sim 100 h^{-1}$ Mpc). While HIPASS detected 1 galaxy every ~ 5 square degrees, ALFALFA is detecting ~ 5 galaxies deg $^{-2}$. The current detection rate suggests that ALFALFA will have detected over 30,000 extragalactic HI line sources when completed. An account of the ALFALFA survey goals and technical characteristics, as well as its main contributions to science, can be found on the website <http://egg.astro.cornell.edu/alfalfa/> and references therein.

This is the first of two papers in which we deal with a compilation of the new ALFALFA data, partially released already in several installments (Giovanelli et al. 2007; Saintonge et al. 2008; Kent et al. 2008; Stierwalt et al. 2009; Martin et al. 2009). The data are from two separate sky strips of the northern Galactic hemisphere covering a total area of 9 hr in right ascension and 16 deg in declination. This portion of the survey is especially inter-

esting because it overlaps almost entirely with the SDSS footprint, thus offering the possibility of cross-correlating observations of the cold-gas component with multi-wavelength optical measurements. This synergy between the surveys will be exploited in this paper and the accompanying one (Toribio et al. 2010, hereafter Paper II) in order to carry out a systematic analysis of the fundamental properties of nearby gaseous galaxies and their interrelations. In this way we hope both to gain a more complete understanding of the structure of the baryonic component in these systems and to set up unbiased benchmarks for the normalcy of their HI content that rely on integral attributes that are easily accessible to observation. While this is not the first attempt to conduct this sort of research (e.g. Haynes & Giovanelli 1984; Roberts & Haynes 1994; Gavazzi et al. 1996; Solanes et al. 1996; Garcia-Appadoo et al. 2009), both the significant increase in the sample size and the high degrees of homogeneity and completeness of the data will allow us to extend the results of previous studies by addressing issues that fell outside their scope.

In this paper, we describe the main statistical properties of the two large catalogs used in our research. We also: describe the steps followed to cross-correlate the radio and optical sources and to calculate their radial velocities (Section 2); explore different characterizations of the environment of ALFALFA detections that may be useful for defining suitable reference samples for the HI content of galaxies (Section 3); and carry out a comparative study of the distribution of the optical structure, colors, and morphologies of the objects included in the selected datasets (Section 4). In Paper II, we will apply strategies of multivariate data analysis to a complete subset of the HI emitters in low density environments defined here. Our aim in that paper will be both to explore inter-variable correlations and to determine the linear combinations of intrinsic properties that best represent these systems. The relationships we derive will define the standards of normalcy for the neutral gas mass content of field galaxies, and thus provide a fundamental reference for future research into the evolutionary effects of the galactic environment.

Unless otherwise stated, to calculate intrinsic quantities we adopt a reduced Hubble constant $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 0.7$.

2. MAIN 21-CM AND OPTICAL CATALOGS

We describe here the primary databases from which we assembled the galaxy samples used in the present study and examine their statistical properties. This section also outlines the steps followed to add information on the radial distance of the selected galaxies to the measurements already available.

2.1. Statistical Properties of Catalogs and Cross-identification of Sources

The main catalog used in this paper contains ALFALFA HI measurements from two regions of the high Galactic latitude sky: $0^{\text{h}}30^{\text{m}} < \text{R.A.}(\text{J2000.0}) < 16^{\text{h}}30^{\text{m}}$, $+4^{\circ} < \text{Decl.}(\text{J2000.0}) < +16^{\circ}$ and $+24^{\circ} < \text{Decl.}(\text{J2000.0}) < +28^{\circ}$. This gives a total coverage of 2160 deg^2 .

The ALFALFA HI detection threshold and completeness limit depend on the source signal-to-noise ratio, S/N , given by the expression (cf. Martin et al. 2009):

$$S/N = \frac{1000 F_{\text{HI}}}{W_{50}^{\text{obs}}} \frac{w_{\text{sno}}^{1/2}}{\sigma_{\text{rms}}}, \quad (1)$$

where F_{HI} is the 21-cm line flux integral in Jy km s^{-1} , which in combination with the cosmological distance to the source, d , in $h^{-1} \text{ Mpc}$ (see Section 2.2), provides its total neutral gas mass in $h^{-2} M_{\odot}$ units

$$M_{\text{HI}} = 2.356 \times 10^5 d^2 F_{\text{HI}}, \quad (2)$$

W_{50}^{obs} is the observed velocity width of the source line profile at the 50% level of the two peaks corrected for instrumental broadening in km s^{-1} , w_{sno} is a smoothing width (either $W_{50}^{\text{obs}}/20$ for $W_{50}^{\text{obs}} < 400 \text{ km s}^{-1}$ or 20 for $W_{50}^{\text{obs}} \geq 400 \text{ km s}^{-1}$), and σ_{rms} is the rms noise figure measured in mJy at a resolution of 10 km s^{-1} . ALFALFA detections are coded according to a scheme dependent on S/N and availability of a prior, e.g., an optical source that matches in angular coordinates or redshift. Code 1 sources are detections with a $S/N > 6.5$ and thought to be reliable with high confidence level; code 2 sources typically have $4 < S/N < 6.5$, but the marginal HI detection is corroborated by a match in position and redshift of an optical source. By searching through the ALFALFA data for code 1 and code 2 sources, we came up with 15,043 extragalactic objects selected purely on their HI content in the sky region defined above and over the full survey bandwidth.

The portion of the ALFALFA survey we use overlaps with other major galaxy surveys such as 2MASS and SDSS. The latter—which has scanned about one quarter of the northern Galactic cap sky in the bands u , g , r , i , and z , reaching objects up to a total magnitude of 22.2 in r band (Stoughton et al. 2002)—is especially interesting for our purposes, as it provides a uniform, very extensive list of photometric and spectroscopic measurements suitable for carrying out an exhaustive study of the correlations between the gaseous and stellar properties of galaxies. Here we make use of SDSS data coming from the full survey (SDSS DR7; Abazajian et al. 2009), which also provides radial velocity measurements for a considerable number of its main targets. The spectroscopic observations, which are complete to a Petrosian r -band magnitude of 17.77 after correcting for Galactic extinction (but not for

internal reddening), offer a redshift coverage that is deep enough to allow the identification of the most probable optical counterpart to all ALFALFA detections found out to ~ 260 Mpc within the sky region where the two surveys overlap¹. A graphical representation of the positions of the ALFALFA sources in the overlapping region is provided in Figure 1 by means of a wedge diagram in right ascension. The total number of HI detections coded 1 and 2 in the SDSS overlap region is 11,239 after discarding objects that lie closer than 1.5 deg to M87, where HI sources cannot be reliably detected because of the strong radio continuum emission.

Some of the main statistical properties of this latter data set are displayed in Figure 2. In the top left panel, we show the distance distribution of the ALFALFA sources. The most important large-scale features are reflected in the various peaks between ~ 80 and 160 Mpc (see also Figure 1). The shape of this histogram follows that produced by the distance distribution of SDSS galaxies (black solid line) closely, except for the narrow artificial dip seen near ~ 125 Mpc and the broader one between ~ 215 and 230 Mpc. Both of these are related to the important loss in sensitivity of the ALFALFA survey due to radio frequency interference (RFI) coming from the San Juan airport’s radar (Giovanelli et al. 2007). The HI mass distributions for the full data set and from code 1 detections only (hatched histogram) are shown in the top right panel of Figure 2. The vertical dotted line illustrates the minimum detectable HI mass arising from the lower heliocentric velocity cutoff of 2000 km s^{-1} adopted in Section 3.1 so as not to include the most uncertain values of this quantity. (Note that this also leaves many nearby dwarfs that are rich in neutral gas out of the present study.) Below, the bottom left panel features the Spänhauer diagram, detailing the HI mass vs. CMB distance. The broad gap in detections associated with the region most affected by the intrusion of RFI is clearly visible near the right-hand end. Finally, the bottom right panel of Figure 2 shows the intrinsic velocity width distribution. As in the top right panel, we also depict the distribution arising only from code 1 objects to demonstrate that the inclusion of code 2 detections does not bias the results.

Unlike other HI surveys, the improved angular resolution of ALFALFA—which leads to positional accuracies typically better than 20 arcsec (Giovanelli et al. 2007)—usually makes the identification of the optical counterpart of a given HI source an unambiguous task. Furthermore, the fact that, for each HI detection, the ALFALFA catalog provides the centroid of the optical galaxy that appears to be the most plausible counterpart according to proximity in z -space, color, and morphology (Giovanelli et al. 2007), enabled us to perform a semi-

¹See http://egg.astro.cornell.edu/alfalfa/images/alfalfa_sdss_coverage.jpg for an image of the footprint of ALFALFA on the SDSS, whose geometry we defined using mangle polygons (Hamilton & Tegmark 2004).

automated catalog cross-correlation. Specifically, we assumed that the optical counterpart of a HI detection provided by the ALFALFA catalog and a spectroscopic SDSS detection are the same object if their projected separation on the plane of the sky is not larger than the size of the optical source and the difference in the reported velocities is less than 300 km s^{-1} . We allow for such a relatively large discrepancy between the observed heliocentric velocities in order to account for the fact that SDSS spectroscopic measurements on an extended object do not always reflect the radial velocity of the object’s centroid, so a positive match between two galaxies relies essentially on the similarity of their sky coordinates. In practice, however, the dispersion in the difference between radial velocities of matched pairs is only of $\sim 35 \text{ km s}^{-1}$ (see Figure 3). For the cases in which ALFALFA sources were lacking a spectroscopic counterpart, the photometric criterion was applied. Human intervention was reduced to only cases of doubtful identification.

Table 1 lists the numbers (and fractions) of ALFALFA objects with a photometric or a spectroscopic counterpart in the SDSS for different bins of radial heliocentric velocity. Only $\sim 2\%$ of the HI sources are not associated with galaxies listed in the SDSS photometric catalog, while there is a somewhat larger percentage, $\sim 15\%$, of HI detections without a spectroscopic counterpart. Note also how the slight incompleteness of the SDSS DR7 data set that affects mainly the galaxies with the largest apparent brightness (e.g., Strauss et al. 2002) leaves its imprint in the form of reduced fractions for the first two velocity bins of the third and fourth columns of the table. Also included in Table 1 are the numbers (and fractions) of ALFALFA objects detected within the ‘SDSS spectroscopic galaxy sample’, SDSS-spec sample for short. From now on, we will use this shorter designation to refer to the subset of SDSS sources with both photometric and spectral morphology classified as of galaxy type, and with $r < 17.77 \text{ mag}$. About one in every five galaxies in this subset is also an ALFALFA detection, although the fraction rises to one in two for $cz_{\text{hel}} \leq 7500 \text{ km s}^{-1}$. The small value quoted in the higher velocity entry of the last column of Table 1 also reflects the impact of the substantial incompleteness of the ALFALFA data due to RFI above $\sim 15,000 \text{ km s}^{-1}$. The effects on the completeness of a few other narrower RFI features present in the spectral bandpass are comparatively modest.

Figure 4 shows how the galaxies in both catalogs are distributed in a global color-light concentration parameter space. The plot shows $(u-r)$ color versus inverse light concentration index, defined as the ratio of the radii containing 50% and 90% of the Petrosian galaxy light, $C_{59} = R_{50}/R_{90}$, which we calculate in the r -band. The density contours delineate the behavior of the galaxies within the SDSS-spec sample, while the data points correspond to those objects that are also ALFALFA members. It is clear from this figure that SDSS galaxies show a clear bimodality and separate into two main classes with comparable numbers of objects that can be addressed as the blue- and red-sequence galaxies (e.g. Blanton et al.

2003). In contrast, HI detections mostly populate the locus corresponding to the bluer, late-type, star-forming systems. This trend has been known about for a very long time (e.g., Roberts & Haynes 1994) and it is in very good agreement with the similar offset found by West et al. (2009) using 195 HI-selected galaxies from the Parkes Equatorial Survey (Garcia-Appadoo et al. 2009). We will show in Section 4.1 that, while the bimodality of the optically-selected sources is reinforced when looking at the global trend with increasing radial velocity, and hence with increasing typical luminosity of the galaxies, the HI detections are distributed in a similar way independently of their distance from the observer.

2.2. Velocity and Radial Distance of Sources Referred to the CMB Frame

To calculate radial velocities and distances referred to the Cosmic Microwave Background (CMB) standard of rest for the galaxies in both the ALFALFA and SDSS spectroscopic data sets, we first correct the heliocentric velocities listed in these two catalogs for peculiar motions. This is done by adopting the multiattractor model of the local peculiar velocity field determined by Masters (2005) from the best fit to data from the new *I*-band Tully-Fisher catalog SFI++ (Springob et al. 2007). This model includes the spherical infall into Virgo and the Hydra-Centaurus supercluster, a quadrupole correction to account for the smaller than average expansion out of the supergalactic plane, as well as a residual dipole pointing towards the Hydra-Centaurus/Shapley superclusters. The predicted peculiar motion corrections, typically of the order of 200 km s^{-1} , are non-negligible for objects within 6000 km s^{-1} . Beyond that radial velocity we assume the Hubble law is exact. The peculiar velocities are tapered to zero between 5000 and 6000 km s^{-1} in order to provide a smooth transition between these two regimes. For transformations between heliocentric and CMB frames of reference, we apply the corrections quoted in Kogut et al. (1993). For conversion to a CMB luminosity distance, we assume Euclidean properties of space, an approximation that is maintained for the calculation of any intrinsic galactic property involved in the present study.

To minimize the impact of redshift-space distortions in the assignment of distances (and, consequently, in local density estimates; see Section 3.1) to galaxies lying in the vicinity of the groups and clusters present in the cosmic volume targeted by ALFALFA, we adopt a strategy that can easily be applied in a systematic way to any virialized aggregation of galaxies independently of its scale. It consists of estimating the virial radius of the system, r_{vir} , and its velocity dispersion along the line-of-sight, σ , and then using this information to assign the corresponding system distance to all galaxies with *projected* distances to the center of one of these overdense environments of less than $1.5 r_{\text{vir}}$ and systemic velocities

with respect to the system mean velocity less than 3σ . To calculate r_{vir} in Mpc, we use the following approximation to the halo density-radius relation based on a singular isothermal sphere with mean interior density 200 times the current value of the critical density of the universe (cf. Carlberg, Yee, & Ellingson 1997):

$$r_{\text{vir}} \simeq \frac{\sqrt{3}\sigma}{700} . \quad (3)$$

Values for σ , in km s^{-1} , mostly come from the compilation of Abell cluster properties by Struble & Rood (1991), while for the few galaxy aggregations whose velocity dispersion is not quoted in that catalog, we use the membership assignments carried out by Springob et al. (2007) in the standard estimator (e.g., Tago et al. 2008)

$$\sigma^2 = \frac{1}{(1 + \bar{z})^2} \sum_{i=1}^{N_v} \frac{(cz_i - c\bar{z})^2}{N_v - 1} , \quad (4)$$

where the sum is intended over the N_v members of the aggregation with measured velocity cz_i , and \bar{z} is the mean system redshift.

3. HI CONTROL SAMPLES

Because the cold gas component of galaxies is both a good indicator of their potential for star formation and an excellent probe of the physical conditions of the intergalactic medium, the quantification of the HI content in galaxies is a useful diagnostic tool for studying the evolutionary effects of their surroundings.

The definition of a benchmark suitable for absolute measures of HI content necessitates a control sample as bias-free as possible formed by galaxies expected to show little or no evidence of environmental interactions. In this section, we study different characterizations of the local environment of ALFALFA detections, define three possible control samples (see Table 2), and explore their relative degree of adequacy.

3.1. Local Number Density Measurements for ALFALFA and SDSS Galaxies

The environment of a galaxy can be defined in terms of the density of galaxies located in its immediate vicinity. To calculate the underlying three-dimensional galaxy number density at a given position, we use the galaxies in the SDSS-spec data set as the nodes of a n th-nearest neighbor network. This calculation is carried out for both ALFALFA and SDSS

sources, excluding galaxies lying in those regions in which the results are affected by large errors or where detections are very scarce and/or unreliable. In particular, we impose a lower heliocentric velocity limit of 2000 km s^{-1} to avoid those galaxies likely having the largest relative uncertainty in distance arising from peculiar motions not duly accounted for—a constraint that also removes the closest cosmic volume where the SDSS pipeline photometry of large, extended galaxies is most inaccurate—and a velocity cutoff ranging from $15,000 \text{ km s}^{-1}$ up to $16,000 \text{ km s}^{-1}$ that embraces the deep trough in sensitivity of the ALFALFA’s spectral window due to RFI (see Figures 1 and 2).

We follow Casertano & Hut (1985) and compute a local, continuous measure of environment from the unbiased estimator

$$\rho_6 = \frac{15}{4\pi d_6^3}, \quad (5)$$

where d_6 is the three-dimensional (comoving) distance to the 6th-nearest neighbor in the SDSS-spec sample in $h^{-1} \text{ Mpc}$. In an effort to correct for redshift-space distortions² and improve the accuracy of our density measures in overdense environments, ρ_6 is calculated, for those galaxies assigned to a cluster (Section 2.2), using instead projected separations scaled by a factor of $(3/2)^{1/2}$ to convert them into a three-dimensional measure of distance (Cooper et al. 2005).

Local densities calculated this way are then corrected for the magnitude cutoff of the SDSS catalog by multiplying them by the inverse of the selection function

$$\psi(\alpha, \delta, z) = \frac{\int_{-\infty}^{M_{\text{lim}}(z) + \Delta M(\alpha, \delta)} \phi(M) dM}{\int_{-\infty}^{M_{\text{fid}}} \phi(M) dM}, \quad (6)$$

which we calculate by integrating the differential luminosity function, $\phi(M)$, of the SDSS provided by Blanton et al. (2001). In equation (6), $M_{\text{lim}}(z)$ and $M_{\text{fid}} \equiv M_{\text{lim}}(z_{\text{fid}})$ are, respectively, the faintest magnitudes detectable because of the optical flux limitation at the galaxy redshift and at a fiducial redshift, z_{fid} , which we match up with that corresponding to the low velocity cutoff of 2000 km s^{-1} imposed to the samples, while $\Delta M(\alpha, \delta)$ is the extinction correction as a function of right ascension (α) and declination (δ), for which we use the Schlegel, Finkbeiner, & Davis (1998) values loaded in the SDSS database.

It is worth noting, however, that this measure of the local density cannot be expanded to the full ALFALFA survey, which is also carrying observations within the solid angle $22^{\text{h}} \leq \text{R.A. (J2000.0)} \leq 03^{\text{h}}$ and $0^\circ \leq \text{Decl. (J2000.0)} \leq +36^\circ$, not scanned by the SDSS. In

²We do not correct for survey edge effects given the large size of the catalogs.

this case, the 2MASS (Skrutskie et al. 2006), with its full sky coverage (98% for the extended source catalog), offers the most sensible alternative to characterizing the environment in which ALFALFA detections reside. Thus, in anticipation of future research, we also assigned a *surface* density parameter μ_6 to each optical and gaseous spectroscopic source from the formula

$$\mu_6 = \frac{5}{\pi s_6^2}, \quad (7)$$

with s_6 the projected distance to the 6th-nearest neighbor, in degrees, derived taking into account the underlying 2MASS galaxy distribution on the sky in the K_s band down to 13.0 mag, where the 2MASS extended source catalog has been estimated to be at least 95% complete for most of the sky (Jarrett et al. 2000). We do not correct this surface density estimator for Galactic absorption, because at this extreme near-infrared wavelength the opacity is insignificant for $|b^{II}| \gtrsim 10^\circ$. It should not be forgotten, however, that this immunity to the effects of the Milky Way’s dust arising from 2MASS being a survey in the near-infrared, biases its sensitivity regarding the morphology of the detected galaxies toward early-type spirals and ellipticals.

In Figure 5, we compare the values of ρ_6 (uncorrected for redshift-space distortions) and μ_6 for galaxies in the SDSS-spec (top) and ALFALFA (bottom) data sets. The plots show that these two local density tracers follow a weak, non-linear relationship, with the three-dimensional measure of environment spanning a somewhat larger dynamic range. Note the relatively few observations associated with clusters in the bottom plot caused by the undersampling of the highest density environments by ALFALFA (see also next section).

3.2. The Low Density Environment H I Galaxy Sample

One possibility to define a sample of galaxies whose evolution is hypothetically less influenced by external, disruptive processes is to select those objects least affected by the clustering phenomenon. According to observational findings the influence of the cluster environment on the properties of galaxies can extend up to about twice the cluster virial radius (e.g., Balogh et al. 1998; Solanes et al. 2001). Therefore, any density threshold adopted to extract a control sample of unperturbed galaxies from a given catalog should be low enough to leave out of it, not only bona fide cluster and group members, but also those objects located in adjacent regions.

We studied the relationship between the neutral gas content of ALFALFA galaxies and their local three-dimensional number density measured in the previous section to see if we can identify the density threshold above which environmental interactions start to become

important. Following the works by Haynes & Giovanelli (1984) and Solanes et al. (1996), the former property has been quantified by means of a HI-deficiency parameter, DEF_{HI} , that compares (the decimal logarithm of) the observed HI mass, $M_{\text{HI}}^{\text{obs}}$, in solar units, with the value expected from a galaxy free of external influences of the same observed morphological type, T^{obs} , and optical linear diameter, D^{obs} , expressed in kpc. Specifically:

$$\text{DEF}_{\text{HI}} = \langle \log M_{\text{HI}}(T^{\text{obs}}, D^{\text{obs}}) \rangle - \log M_{\text{HI}}^{\text{obs}}, \quad (8)$$

so positive values of DEF_{HI} indicate HI deficiency. This test has been restricted to those ALFALFA detections with Hubble types and (apparent) blue visual sizes available on the Arecibo General Catalog (AGC), a private galaxy database maintained by MPH and RG. In order to increase sample size and reduce statistical errors, we inferred expectation values for the HI mass for 2615 disk galaxies with Sa–Sd morphologies using the maximum likelihood linear regressions of $\log M_{\text{HI}}$ on $\log D^{\text{obs}}$ given in Solanes et al. (1996). Since in that work the standards of normalcy for the HI content were defined specifically for galaxies of type Sa–Sc, we thus assigned Scd–Sd types to be the same as Sc.

We present in Figure 6 a plot of the median DEF_{HI} in bins of local number density. We show results inferred using sources with radial velocities between 2000 and 18,000 km s^{−1}, as well as separately for three sub-intervals of this velocity range. Looking at the different panels one can see moderate oscillations in the central values of DEF_{HI} . Nonetheless, these are always well within the quoted uncertainties and therefore not statistically significant, hinting at best a mild increase of gas deficiency with density. The weak dependence of the gas content of ALFALFA sources with the local density is the consequence of limiting the effective integration time of the survey to about 48 seconds per beam area. While ALFALFA has been designed to detect from massive spirals at $z \sim 0.06$ (~ 250 Mpc) to dwarfs with HI masses as low as $M_{\text{HI}} \sim 10^7 M_{\odot}$ at the Virgo distance (e.g., di Serego Alighieri et al. 2007), the highly gas-deficient galaxies that are expected to reside in the densest regions of the surveyed volume should escape detection unless they are located nearby (see also Figure 5). It is also important to stress that the observed behavior of the gas content with local density is not affected by the negative offset shown by the inferred median DEF_{HI} values. Such a bias obeys to several factors. On one hand, as mentioned above, we are underestimating the expected gaseous content of Scd–Sd galaxy types, which lowers the median DEF_{HI} of our dataset about 0.15 units. Moreover, it is the fact that ALFALFA is detecting total 21-cm fluxes for individual galaxies that, on average, are around 10% larger than their AGC counterparts. This reduces the average value of DEF_{HI} by additional 0.05 units. The rest of the drop in the HI deficiency can be attributed to differences in the object selection with respect to Solanes et al. (1996). Note that the ALFALFA survey extends further in redshift than the sample used in this former work, which was centered on the Pisces-Perseus Supercluster. This allows us to deal with a significantly large number of high-HI-mass

objects. Besides, due to a less robust selection of the control sample, we cannot discard the presence in the Solanes et al.’s data of a non-negligible fraction of galaxies with a disturbed HI mass, which might have contributed to further reduce the standard values of this quantity calculated in that paper.

The lack of a clear HI content-density correlation for the portion of the ALFALFA spring-sky survey under study notwithstanding, we prefer to be conservative and, as indicated at the beginning of the section, define a control sample that does not include galaxies located in the vicinity of an overdensity. After testing different density thresholds, we find that the condition $\rho_6 \leq 0.5 h^3 \text{ galaxies Mpc}^{-3}$ discards all SDSS galaxies with spectroscopic information cataloged as cluster members in the surveyed region —the corresponding threshold based on the 2MASS galaxy distribution would be $\mu_6 \sim 10 \text{ galaxies deg}^{-2}$ —, as well as more than 91 (96) per cent of these objects with system-centric distances between $1.5\text{--}3 r_{\text{vir}}$ ($1.5\text{--}2 r_{\text{vir}}$) and systemic velocities within 1σ of the cluster mean, which may be considered representative of the system outskirts population. This threshold performs similarly well applied to ALFALFA detections. The Low Density Environment (hereafter LDE) HI galaxy sample that results from selecting those ALFALFA detections obeying the above density constraint has a total of 5647 objects. Wedge diagrams in right ascension of the spatial distribution of this data set, as well as the HI detections discarded for lying in higher density environments, are shown in Figure 7.

3.3. Isolated Galaxy Samples

By selecting galaxies in low density regimes one may confidently exclude objects associated with clusters and large groups, but cannot guarantee that such galaxies have not interacted at some point with nearby companions. That is the reason why we also studied the possibility of defining control samples from the ALFALFA catalog that maximize the chances of dealing with galaxies whose evolution has not been influenced by other objects in the recent past. Here we build two samples of isolated galaxies by applying two different kinds of isolation criteria.

As a first approach, we devised a selection technique for the SDSS spectroscopic data set, similar to the one used by Prada et al. (2003), that combines spectroscopic with photometric information to detect galaxies without *dynamically relevant* companions inside a given volume in z -space. Thus, we classify a galaxy i with apparent r -band total magnitude r_i at coordinates $(\alpha_i, \delta_i, z_i)$ as a good candidate for a sample of isolated galaxies if it has no neighbor j with $r_j \leq r_i + 2.5 \text{ mag}$ within the cylindrical volume, centered on its position, of

radius and semi-axis

$$R = R_0\psi^{-1/3} \quad \text{and} \quad V = V_0\psi^{-1/3} + \sigma_{12} , \quad (9)$$

respectively, where $R_0 = 0.7$ Mpc is the linking-length at the adopted fiducial redshift of 2000 km s⁻¹, $V_0 = H_0 R_0$, and $\psi(\alpha_i, \delta_i, z_i)$ is the selection function defined in equation (6) that keeps constant the probability of finding a galaxy in a given volume regardless of distance. In equation (9), the term $\sigma_{12} = 350$ km s⁻¹ is added in order to account for the distortion on the intergalactic distances along the line-of-sight induced by the one-dimensional pairwise velocity dispersion of field galaxies (Guzzo et al. 1997; Landy 2002). The preference of a 0.7 Mpc linking-length has been dictated from the fact that it would take almost 3 Gyr for a galaxy with a peculiar velocity of 175 km s⁻¹ to cover such a distance, while the magnitude constraint is set to take into account neighbors that have at least one-tenth of the target mass (assuming light is a proxy for mass). This technique identifies 461 SDSS-spec galaxies that are also HI detections as members of this first isolated galaxy (hereafter IG1) sample.

A second isolated galaxy (IG2) sample has been defined by applying the sort of prescription typically used for photometric samples —usually much deeper than the spectroscopic ones—, which takes into account in a semi-empirical way the distance to companions. Perhaps the best known example is the nearest neighbor algorithm by Karachentseva (1973) based on the comparison of apparent angular diameters between candidate isolated galaxies and their neighbors. Recently, Allam et al. (2005) have implemented a variation of Karachentseva’s technique to detect isolated galaxies in the SDSS DR1 that we adopt here with a slight modification. Thus, we consider a target galaxy i with a g -band total magnitude g_i and g -band Petrosian radius R_i to be isolated if the projected sky separation $\theta_{i,j}$ with respect to any neighboring galaxy j satisfies the conditions

$$\theta_{i,j} \geq 40R_j \quad \text{and} \quad |g_i - g_j| \leq 2.5 , \quad (10)$$

where the 2.5 magnitude difference corresponds to about a factor 9 in brightness for a galaxy with a flat surface brightness profile. We also took into account the fact outlined by Allam et al. (2005) that the application of these selection criteria to SDSS data effectively reduces the magnitude range of the candidate isolated galaxies at both ends (see, for instance, their Figures 11 and 12). Thus, by considering only candidates with Galactic-extinction-corrected g -band magnitudes in the range $15.5 \leq g_i \leq 19.5$, we extracted a total of 236 isolated HI emitters³.

³Had we followed the same approach, but imposing instead the more restrictive conditions $|g_i - g_j| \leq 3.0$ mag with $16.0 \leq g_i \leq 19.0$ mag used in Allam et al.’s paper, the number of retrieved isolated HI galaxies would have dropped to 67 within the region surveyed by ALFALFA.

Given that the spectro-photometric algorithm imposes the restriction $r_i \leq 15.27$ on the magnitude of IG1 members, there is only a small overlap in the range of apparent brightnesses of the galaxies that are included in both subsets of isolated objects. In practice, this means that only 7 ALFALFA detections end up obeying simultaneously the two isolation criteria. If we eliminate from the photometric prescription the constraint $g_i \geq 15.5$ mag to make the overlap of both subsets of isolated galaxies easier, then the number of IG2 members rises to 470. Even so, this enlarged sample includes only 83 of the 461 isolated galaxies selected by the spectro-photometric algorithm. Therefore, isolated galaxies in z -space are not necessarily considered so when their environment is examined in projection and vice versa.

As done in the previous section for the LDE HI galaxy sample, we also studied the environment of the ALFALFA galaxies in the IG1 and IG2 subsets. As an example, we display in Figure 8 a graphical representation of the HI-deficiency versus local density for IG1 galaxies. As before, we focus only on sources having optical diameters and spiral types from Sa to Sd listed on the AGC. As shown in this figure, while isolated galaxies obtained through automated systematic searches avoid the densest regions of the universe, a non-negligible fraction of them inhabit regions of moderate local galaxy density having values of ρ_6 larger than the upper density threshold adopted to define our LDE HI galaxy sample. This demonstrates that the definitions of isolation and low density environment are not necessarily equivalent; local number density measurements associated with isolated galaxies can span quite a broad range of values (for a similar conclusion, see Verley et al. 2007). Alongside this fact, Figure 8 also shows that although the different approaches adopted to identify field galaxies assemble samples of significantly different sizes, the corresponding gas content distributions are quite similar. There is therefore no evidence that our searches of isolated galaxies selected objects that are particularly gas-rich compared to those in the LDE galaxy sample.

4. UNDERLYING PROPERTIES OF GALAXIES IN THE HI CONTROL SAMPLES

Here we use several independent classification schemes based on the photometric data for the galaxies in the three candidate control samples defined in the previous section to provide more details and increase our understanding of the objects that inhabit quiet environments. These measures are all sensitive to the morphology and they are objective, easily reproducible, and applicable to data sets in which the morphological classification by direct inspection of the galaxy images is impractical, given the large number of objects involved.

4.1. Structure, Colors, and Morphology of the LDE Galaxies

Since we are focusing our attention on subsets of non-clustered, gas-rich galaxies, predictably of late type, we first apply the criterion used by Maller et al. (2009; hereafter Mal09) to select disk galaxies from the SDSS. As stated by those authors, almost pure disk samples can be obtained by selecting galaxies that obey *at least one* of these requirements: Sérsic index $n \leq 3$ or observed axial ratio $b/a \leq 0.55$, where the parameter n —which in our case has been extracted from the NYU Value-Added Galaxy Catalog (Blanton et al. 2005)—measures the shape of the observed r -band luminosity profile of a galaxy fitted using the Sérsic $R^{1/n}$ formula with elliptical isophotes. We note that these constraints include S0 galaxies in the ‘disk’ class, because the presence of a disk makes their measured axial ratio more important in determining their apparent inclination than the intrinsic ellipticity of the spheroidal component. According to Mal09, the completeness of their classification scheme reaches 70% for disk galaxies, with a reliability of $\sim 95\%$ when applied to SDSS data. The completeness is the fraction of galaxies of a given type that are successfully selected from the original sample by the classification scheme. The reliability is the fraction of galaxies of the desired type in the selected subset.

In good agreement with expectations, Table 3 shows that the fraction of members of the SDSS-spec sample in low density environments ($\rho_6 \leq 0.5 h^3$ galaxies Mpc^{-3}) that obey Mal09’s criterion is very high ($\sim 85\%$ on average), and even higher for the subset of ALFALFA detections ($\sim 92\%$). Our global percentages of late-type objects are, however, slightly lower than those obtained by Mal09. This result can be attributed to the fact that they were inferred for a very shallow data set with $M_r < -20.6$ mag (note that the agreement between our results and Mal09’s becomes almost perfect if we restrict the comparison to our lowest velocity bin). In principle, the decrement in the fraction of galaxies assigned to the disk class with increasing radial velocity illustrated in Table 3 can be interpreted as an indication of the fact that most luminous galaxies tend to be of early type, in qualitative accordance with previous work (e.g., Strateva et al. 2001; Blanton et al. 2003; Baldry et al. 2004). Yet we find that in the highest velocity bin $\sim 10\%$ of the ALFALFA galaxies are not classified as disks, while at these relatively large distances HI detections should arise almost exclusively from late-type objects: at $11,000 \text{ km s}^{-1}$ the minimum HI mass detectable in the survey is $\log(M_{\text{HI,min}}) \sim 9.3 M_\odot$. Accordingly, our results seemingly highlight the difficulties in morphology assignment that selection rules like Mal09’s, based on global measures of galaxy structure, encounter for faint objects due to the inherent noisiness of the images.

The inverse light concentration index, $C_{59} = R_{50}/R_{90}$, is another measure of galaxy structure that can be used as a morphological divider (Shimasaku et al. 2001; Strateva et al. 2001). As with the Sérsic index, this parameter is expected to be correlated with galactic

morphology. Indeed, the two indexes can be interchanged for an ideal galaxy with a Sérsic profile in the absence of seeing. Early-type galaxies have a dominant bulge component, so their light is concentrated, while late-type galaxies, which harbour bulges of varying size, have their light more dispersed due to the presence of an extended exponential disk component and hence larger C_{59} values. In practice, however, the concentration index is a morphological separator that is useful for broadly dividing galaxies into two populations: early-type (E, S0, Sa) and late-type (Sb, Sc, Irr) objects. We follow Strateva et al. (2001) and adopt an $C_{59,r} = 0.38$ separator in the r -band, which provides a cut giving equally complete subsamples of both morphological populations, while at the same time it reduces the contamination of the late-type galaxy class by early-type objects to $\sim 12\%$. When applied to galaxies in LDEs, we find that most galaxies have $C_{59,r} > 0.38$, indicating the dominance of disks. As was to be expected, this tendency is again more marked for ALFALFA detections, about 82% of them fall into the light-concentration-index late-type class, than for the optical subset, which has a late-type galaxy fraction of $\sim 70\%$.

Given the correlation between color and morphology, early-type galaxies being generally redder than late-types, galaxy colors can also be used as a morphological classification tool, as well as to provide a coarse separation of galaxies in terms of their star formation activity. We adopt the classification scheme of Baldry et al. (2004), which uses the SDSS model magnitudes (Stoughton et al. 2002) to measure color from the difference between the fluxes in the u and r filters and to calculate the absolute magnitude in the r band, M_r , for which we adopt the CMB distance inferred in Section 2.2. As well as being an optimal color for separating galaxies into red and blue classes (Strateva et al. 2001), $(u - r)$ is a good indicator of star formation activity (Deng & Zou 2009), while the r filter contains the peak of the light curve for most SDSS galaxies. In this analysis, we do not apply K -corrections since the samples studied are shallow.

Figure 9 shows the color-magnitude diagram for ALFALFA detections in the LDE sample with the Baldry et al. delimiter,

$$(u - r) = 2.06 - 0.244 \tanh \left(\frac{M_r + 20.07}{1.09} \right), \quad (11)$$

superposed. Not surprisingly, we see quite a few more sources on the blue, active population side of the divider (below the curve) than on the red, passive side: 83% and 17% of the total, respectively. Moreover, we note that most of the H I galaxies above the separator reside in an area of the color-magnitude diagram around $M_r \lesssim -20$ mag and $(u - r) \gtrsim 2.3$ mag where the blue and red galaxy populations identified by Baldry et al. (2004) overlap in this space (see their Figure 9). Therefore, it is not unlikely that many of these galaxies are actually blue interlopers. Indeed, if we repeat the same exercise after correcting the SDSS magnitudes

and colors to face-on values, following respectively Shao et al. (2007) and Masters et al. (2010a), and keeping the same separator, the fraction of HI detections assigned to the blue group increases to 92%. This highlights the importance of defining morphological separators corrected for internal extinction, since the effects of dust on the colors and concentration of late-type galaxies can be significant for edge-on systems.

A slightly coarser color division, but one that facilitates the analysis of the data, is provided by the luminosity-independent ($u - r$) color cut of 2.22 mag —also uncorrected for internal extinction— suggested by Strateva et al. (2001). In the four panels of Figure 10, we show the observed color distributions of the SDSS-spec galaxies in LDEs separated into HI-detections and non-detections for different bins of radial heliocentric velocity. As illustrated by the histograms, the global fraction of SDSS galaxies in LDEs that are of red-/early-type becomes notably higher with increasing distance, in agreement with the well-known color-luminosity-morphology(-star formation activity) correlation (see also Figure 9), which appears to hold regardless of local galaxy density (e.g., Ball, Loveday, & Brunner 2008). Yet it is obvious from these plots that the observed behavior arises almost exclusively from the contribution of galaxies undetected in HI. For these systems, the fraction of red objects, according to the Strateva et al. (2001) separator, grows steadily from less than 15% in the lowest velocity bin, to about 39% in the highest one, where $\sim 96\%$ of the SDSS galaxies redder than the divider are HI non-detections. In sharp contrast, field HI emitters consist of (mostly blue) objects whose color distribution shifts only very slightly redwards with increasing distance, to the point that the ratio between the blue and red fractions for this subset remains nearly constant at $\sim 11 \pm 4\%$ in all redshift bins ($\sim 5 \pm 2\%$ if one uses colors corrected for internal extinction) and, hence, approximately independent of the mean HI mass of the detections. This result contrasts with previous claims, based on optically-selected sources, that blue exponential systems, just as with red concentrated ones, become monotonically redder with increasing luminosity (e.g. Blanton et al. 2003; Ball et al. 2008) and suggests that the color and HI mass of gaseous galaxies cannot be very strongly related (see also Section 4.2). In Paper II we will return to this issue, which also has fundamental implications for the theory of galaxy formation regarding the number of independent physical parameters that determine the properties of individual galaxies (Gavazzi et al. 1996; Disney et al. 2008; Gavazzi 2009).

The Galaxy Zoo 1 (GZ1) clean catalog (Lintott et al. 2010), which provides visual morphological classifications for nearly 900,000 SDSS galaxies, offers the possibility of directly inspecting the basic morphology of a highly significant fraction of our HI detections. To this end, we focused our attention on the subset of 4599 GZ1 objects that are members of the SDSS DR7 spectroscopic survey and are also included in our LDE HI galaxy sample. Following Lintott et al., we were rather strict in assigning GZ1 morphologies to our

ALFALFA objects by requiring that they have a ‘spiral’ or an ‘elliptical’ flag in the clean catalog, i.e., with more than 80% confidence in category assignation after debiasing corrections. All other galaxies were classified as of uncertain type —as shown in Figure 11, most of the ALFALFA detections with uncertain GZ1 morphology are objects with $n < 2$ but $b/a \gtrsim 0.4$. Comparison with the Hubble types listed in the AGC indicates that the reliability of this classification scheme for the spiral class is higher than 90%, with a $\sim 3\%$ contamination from S0/S0a galaxies, in good agreement with estimates by Lintott et al. (2008) and Bamford et al. (2009). In this manner, we infer that $\sim 52\%$ of our HI detections can confidently be assigned a spiral (disk with spiral arms) morphology, whereas only $\sim 1\%$ are galaxies with a high probability of belonging to the elliptical class (E and S0 galaxies). The fact that all the objects in the latter class have $b/a > 0.5$ and that most of them have $n > 3$ (r -band values) implies that prescriptions like that used by Mal09 are very effective at separating out objects of early-type morphology. Indeed, we find that only 19 of the 4057 Mal09 disks identified in this subset of GZ1 members are classified by the Galaxy Zoo as bona fide ellipticals. We note in addition that just 78 of the 4599 members of the LDE-GZ1 HI subset have a weighted-merger-vote fraction larger than 40% (Darg et al. 2010).

We extended this a step further by examining by eye the SDSS images of those LDE gaseous galaxies that show optical early-type characteristics according to both their color and Sérsic index. We concentrated our efforts on objects with extreme inclinations that meet the following criteria: $(u - r) > 2.22$ mag, $n > 4.0$, and $b/a < 0.5$ or $b/a > 0.85$ (the latter two parameters measured in the r -band). In agreement with Mal09, we found that most of our edge-on, red, concentrated HI sources have a clear disk-like morphology suggesting that they owe their early-type classification to large amounts of dust reddening. This morphology also appears to be the dominant one among the face-on gas-rich galaxies identified with large Sérsic indexes and red colors, in which luminous red central regions and large bars are plentiful. We speculate that these latter objects represent a stage of spiral evolution that is not as advanced as that assigned to the population of red spirals recently studied by Masters et al. (2010b), since ours still have very blue, but faint, outer disks and are quite rich in neutral gas (several of them have $M_{\text{HI}} > 10^{10} M_{\odot}$), in good agreement with the findings of the GALEX Arecibo SDSS Survey (Wang et al. 2010). Therefore, we are left with only a handful of cases that may correspond to genuine gas-rich early-type galaxies (ETGs) in which the cold gas could have a recent external origin. While the scarcity of this sort of ‘hybrid system’ in our LDE sample is consistent with the very few large ETGs detected in HI in the Virgo cluster region (di Serego Alighieri et al. 2007), it seems to be at odds with the preliminary results of the ATLAS^{3D} survey of 262 nearby galaxies that lack spiral arms in optical images (Serra et al. 2010), which hint at the possibility that isolated ETGs embedded in massive HI disks are relatively abundant.

The results discussed above, plus the fact that the automated classification of galaxies into two major populations using extensive stellar parameters leads to subdivisions that are partially contaminated by members of the other class, and therefore to the standardization of the trends, lead us to conclude that HI emission offers the most reliable (and clear) way to determine whether a galaxy has genuinely an early-type or late-type morphology.

4.2. Structure, Colors, and Morphology of the Members of the IG1 and IG2 Subsets

The exercises reported in the previous section for ALFALFA detections and non-detections in LDEs were repeated for the samples of isolated galaxies.

As illustrated in Figure 12, we find that, regardless of the isolation criterion adopted, the population of HI detections is notably bluer than that of non-detections, in good agreement with our results for galaxies in LDEs. There are, however, significant differences in the color distributions of both samples of isolated galaxies. Thus, while the colors of the IG2 members mimic the behavior shown by the LDE galaxies at the median depth of the ALFALFA survey (bottom-left panel of Figure 10), the IG1 members display nearly unimodal distributions — there is a hint of bimodality in the histogram of HI detections— strongly skewed to the red. Likewise, we find that IG1 galaxies tend to have both smaller light concentration indexes and a lower fraction of Mal09 disks than galaxies in both the IG2 and LDE data sets, as would be expected from a population more dominated by early-type objects. Indeed, if we restrict our attention to just the ALFALFA galaxies with Hubble types listed in the AGC, we find that the latter two samples are considerably richer in Scd and Sd spiral subtypes than the IG1 subset is.

These discrepancies, however, do not stem from fundamental differences between the properties of the objects belonging to the IG1 and IG2 samples, but from the distinct selection criteria adopted to draw up their member galaxies. As illustrated by the top panels of Figure 13, while the luminosity distribution of the LDE galaxies (as well as that of the full ALFALFA catalog) moves clearly towards brighter values with increasing distance, the corresponding color distribution, which can be reasonably approximated by a unimodal function peaking around the characteristic color of the blue sequence and having a long red tail (see also Figure 10), shows a modest variation. This rough distance-invariance is reinforced by the lack of a substantial red-sequence component for the HI emitters, which start contributing significantly to the color distribution of the entire galactic population at somewhat larger redshifts than the blue one. On the other hand, the isolation selection techniques we used place limits on the apparent brightness that the candidate galaxies can have. This

inevitably leads to choosing from among the brightest objects that are within the allowed magnitude range and, consequently, also among the reddest. Such a bias is stronger in the case of our spectro-photometric IG1 sample, which excludes objects fainter than $r = 15.27$ mag. As a result, most of the HI sources around the peak of the blue sequence are missing from the color distribution of this subset, which only preserves its red tail (middle right panel of Figure 13). In contrast, the almost disjoint constraint $15.5 \leq g_i \leq 19.5$ mag adopted for the definition of the IG2 sample from the relatively shallow ALFALFA data set has led us to select galaxies that are just slightly intrinsically brighter and, therefore, slightly redder too, than the average LDE members (bottom panels of Figure 13). Note, however, that the application of the photometric isolation prescription by Allam et al. (2005) to the much deeper SDSS DR1 catalog led these authors to obtain an isolated galaxy sample with a light concentration index distribution that suggested a fifty-fifty morphological composition between early and late galaxy types, closer to what we observe here for the IG1 subset.

5. SUMMARY AND CONCLUSIONS

We have used measurements in the 21-cm emission line from the ALFALFA blind survey in a region of the sky also scanned by the SDSS DR7 in order to study the physical properties of gas-rich galaxies expected to show little or no evidence of interaction with their surroundings. The goal has been twofold: 1) to improve the understanding of the nature of extragalactic HI sources; and 2) to identify an observational sample of gaseous galaxies as unperturbed as possible by the environment and therefore suitable to set up standards for their neutral gas content.

Different parametric estimates of the environment of HI sources have been explored, in order to establish their suitability in providing a catalog of gaseous systems whose properties are minimally affected by external influences. Among the environmental measures tested, the local number density estimate based on the distance to the 6th-nearest spectroscopic SDSS neighbor of a galaxy has emerged as the most appropriate. By applying this estimator to the 11,239 HI detections identified in the overlapping region of the radio and optical surveys, we have extracted a subset of 5647 HI sources in low density environments. We have also defined two different subsets of isolated galaxies by applying standard isolation criteria based either on the combination of spectroscopic and photometric information or solely on photometric data. In spite of the marked differences in size and membership between the LDE and isolated subsets, their respective HI content distributions have proven to be all pretty similar.

In combination with SDSS data, we have studied the distributions of light concentration

indexes, colors, and other proxies of morphology for the members of these three control samples. All the evidence analyzed indicate that HI emission is detected essentially in objects that are structurally similar to galaxies with late-type optical characteristics. Thus, over 90% of HI sources residing in a quiet environment obey the disk selection rule by Mal09, while some 82 per cent have inverse light concentration indexes representative of this class. These fractions could be considered actually a lower limit, as the determination of the structural parameters of distant objects is quite imprecise. Likewise, from evidence gathered by studying the color distributions, we have found that at least between 85% and 90% of the ALFALFA detections outside overdense regions are classified as blue, star-forming galaxies. These fractions are even higher when extinction-corrected colors are used. In contrast, the structural properties and colors of HI non-detections are typical of a population substantially richer in early-type, red objects. The assignment of morphologies to a subset of ALFALFA galaxies in LDEs from the GZ1 catalog corroborates these conclusions, as only a very small fraction (~ 1 per cent) of the objects studied have been classified into the elliptical class with high reliability. These results suggest that the more reliable automated way to determine whether a galaxy has genuinely an early- or late-type morphology is by measuring its HI emission.

The standard view that red galaxies and, consequently, passive objects of type S0/Sa or earlier, tend to have less neutral hydrogen than bluer, star-forming late-type systems is therefore confirmed. Our research also corroborates the prediction, based on the positive tight correlation between global color and intrinsic luminosity inferred from optically-selected galaxies, that the fraction of red-/early-type galaxies in our SDSS-spec LDE sample should increase with increasing radial velocity of the sources. We observe, however, that while the color distribution of the objects undetected in HI clearly becomes redder, that of the gas-rich population does not change significantly as a function of redshift. This suggests that the color and gaseous mass of field HI emitters are not very strongly interrelated. Likewise, we have shown that the close connection between color and luminosity can have an important impact on isolation criteria, favoring the selection of the redder/earlier members of a catalog, the shallower the survey the most notorious the bias. Awareness of this selection effect, together with the relatively modest sizes and reduced dynamic range subtended by the properties of the members of our isolated galaxy samples—which are a direct consequence of the severe restrictions imposed by the selection algorithms on the range of apparent magnitudes—, leads us to conclude that the LDE approach provides the most suitable and statistically sounder control sample to set standards of reference for the HI content of galaxies. In Paper II we will present a detailed derivation of such standards by applying a non-parametric principal component analysis to a flux-limited subset of this latter dataset.

This work was supported by the Dirección General de Investigación Científica y Técnica, under contract AYA2007-60366. M.C.T. acknowledges support through a fellowship from the Spanish Ministerio de Educación y Ciencia. RG and MPH receive support from NSF grant AST-0607007 and from a grant from the Brinson Foundation. KLM acknowledges funding from the Peter and Patricia Gruber Foundation as the 2008 IAU Fellow, and from the University of Portsmouth and SEPnet (www.sepnet.ac.uk).

We thank the many members of the ALFALFA team who have contributed to the acquisition and processing of the ALFALFA data set over the last six years.

This research is based mainly on observations collected at Arecibo Observatory. The Arecibo Observatory is part of the National Astronomy and Ionosphere Center, which is operated by Cornell University under a cooperative agreement with the National Science Foundation.

Some of the results quoted in this publication have been made possible by the participation of more than 160,000 volunteers in the Galaxy Zoo project. Their contributions are individually acknowledged at <http://www.galaxyzoo.org/Volunteers.aspx>.

We have also made use of the HyperLeda database (<http://leda.univ-lyon1.fr>) and the NASA/IPAC Extragalactic Database, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract from the National Aeronautics and Space Administration. Likewise, we are grateful to all the people and institutions that have made possible the NYU Value-Added Galaxy Catalog (<http://sdss.physics.nyu.edu/vagc/>) and the Sloan Digital Sky Survey (SDSS; <http://www.sdss.org/>). Funding for the SDSS is provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England.

REFERENCES

- Abazajian, K. N., et al. 2009, *ApJS*, 182, 543
- Allam, S. S., Tucker, D. L., Lee, B. C., & Smith, J. A. 2005, *AJ*, 129, 2062
- Baldry, I. K., Glazebrook, K., Brinkmann, J., Ivezić, Ž., Lupton, R. H., Nichol, R. C., & Szalay, A. S. 2004, *ApJ*, 600, 681
- Ball, N. M., Loveday, J., & Brunner, R. J. 2008, *MNRAS*, 383, 907
- Balogh, M. L., Schade, D., Morris, S. L., Yee, H. K. C., Carlberg, R. G., & Ellingson, E. 1998, *ApJ*, 504, L75
- Bamford, S. P., et al. 2009, *MNRAS*, 393, 1324

- Blanton, M. R., et al. 2001, *AJ*, 121, 2358
- Blanton, M. R., et al. 2003, *ApJ*, 594, 186
- Blanton, M. R., et al. 2005, *AJ*, 129, 2562
- Carlberg, R. G., Yee, H. K. C., & Ellingson, E. 1997, *ApJ*, 478, 462
- Casertano, S., & Hut, P. 1985, *ApJ*, 298, 80
- Conselice, C. J. 2006, *MNRAS*, 373, 1389
- Cooper, M. C., Jeffrey, A. N., Madgwick, D. S., Gerke, B. F., Renbin, Y., & Davis, M. 2005, *ApJ*, 634, 833
- Darg, D. W., et al. 2010, *MNRAS*, 401, 1043
- Deng, X.-F., & Zou, S.-Y. 2009, *Astroparticle Physics*, 32, 129
- di Serego Alighieri, S., et al. 2007, *A&A*, 474, 851
- Disney, M. J., Romano, J. D., GarciaAppadoo, D. A., West, A. A., Dalcanton, J. J., & Cortese, L., 2008, *Nature*, 455, 1082
- Garcia-Appadoo, D. A., West, A. A., Dalcanton, J. J., Cortese, L., & Disney, M. J. 2009, *MNRAS*, 394, 340
- Gavazzi, G., Pierini, D., & Boselli, A. 1996, *A&A*, 312, 397.
- Gavazzi, G. 2009, in *A Long Walk Through Astronomy: A Celebration of Luis Carrasco's 60th Birthday*, ed. E. Recillas, A. Luna, & Y. D. Mayya, *RevMexAA Conference Series*, 37, 72
- Giovanelli, R., & Haynes, M. P. 1985, *ApJ*, 292, 404
- Giovanelli, R., et al. 2005, *AJ*, 130, 2598
- Giovanelli, R., et al. 2007, *AJ*, 133, 2569
- Guzzo, L., Strauss, M. A., Fisher, K. B., Giovanelli, R., & Haynes, M. P. 1997, *ApJ*, 489, 37
- Hamilton, A. J. S., & Tegmark, M. 2004, *MNRAS*, 349, 115
- Haynes, M. P., & Giovanelli, R. 1984, *AJ*, 89, 758
- Hubble, E. 1926, *ApJ*, 64, 321

- Jarrett, T. H., Chester, T., Cutri, R., Schneider, S., Rosenberg, J., Huchra, J. P., & Mader, J. 2000, *AJ*, 120, 298
- Karachentseva, V. E. 1973, *Comm. Spec. Astrophys. Obs.*, 8, 3
- Kent, B. R., Giovanelli, R., Haynes, M. P., Martin, A. M., Saintonge, A., Stierwalt, S., Balonek, T. J., Brosch, N., & Koopmann, R. A. 2008, *AJ*, 136, 713
- Kogut, A., et al. 1993, *ApJ*, 419, 1
- Landy, S. D. 2002, *ApJ*, 567, L1
- Lang, R. H., et al. 2003, *MNRAS*, 342, 738
- Lintott, C. J., et al. 2008, *MNRAS*, 389, 1179
- Lintott, C. J., et al. 2010, *MNRAS*, 410, 166
- Maller, A. H., Berlind, A. A., Blanton, M. R., & Hogg, D. W. 2009, *ApJ*, 691, 394 (Mal09)
- Martin, A. M., Giovanelli, R., Haynes, M. P., Saintonge, A., Hoffman, G. L., Kent, B. R., & Stierwalt, S. 2009, *ApJS*, 183, 214
- Masters, K. L. 2005, Ph.D. thesis, Cornell University
- Masters, K. L., et al. 2010a, *MNRAS*, 404, 792
- Masters, K. L., et al. 2010b, *MNRAS*, 405, 783
- Meyer, M. J., et al. 2004, *MNRAS*, 350, 1195
- Prada, F., et al. 2003, *ApJ*, 598, 260
- Ramella, M., Geller, M. J., & Huchra, J. P. 1992, *ApJ*, 384, 404
- Roberts, M. S., & Haynes, M. P. 1994, *ARA&A*, 32, 115
- Rosenberg, J. L., & Schneider, S. E. 2002, *ApJ*, 567, 247
- Saintonge, A. 2007, *AJ*, 133, 2087
- Saintonge, A., Giovanelli, R., Haynes, M. P., Hoffman, G. L., Kent, B. R., Martin, A. M., Stierwalt, S., & Brosch, N. 2008, *AJ*, 135, 588
- Shao, Z., Xiao, Q., Shen, S., Mo, H. J., Xia, X., & Deng, Z. 2007, *ApJ*, 659, 1159

- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525
- Schneider, D. P., et al. 2010, *AJ*, 139, 2360
- Serra, P., et al. 2010, in *Galaxies in Isolation: Exploring Nature versus Nurture*, ed. L. Verdes-Montenegro, A. del Olmo, and J. W. Sulentic, *ASP Conference Series*, 421, 49
- Shimasaku, K., et al. 2001, *AJ*, 122, 1238
- Skrutskie, M. F., et al. 2006, *AJ*, 131, 1163
- Springob, C. M., Masters, K. L., Haynes, M. P., Giovanelli, R., & Marinoni, C. 2007, *ApJS*, 172, 599
- Solanes, J. M., Giovanelli, R., & Haynes, M. P. 1996, *ApJ*, 461, 60
- Solanes, J. M., Manrique, A., González-Casado, G., García-Gómez, C., Giovanelli, R., & Haynes, M. P. 2001, *ApJ*, 548, 97
- Spearman, C. 1904, *Am. J. Psychol.*, 15, 72
- Stierwalt, S., Haynes, M. P., Giovanelli, R., Kent, B. R., Martin, A. M., Saintonge, A., Karachentsev, I. D., & Karachentseva, V. E. 2009, *AJ*, 138, 338
- Stoughton, C., et al. 2002, *AJ*, 123, 485
- Strateva, I., et al. 2001, *AJ*, 122, 1861
- Strauss, M. A., et al. 2002, *AJ*, 124, 1810
- Struble, M. F., & Rood, H. J. 1991, *ApJS*, 77, 363
- Tago, E., Einasto, J., Saar, E., Tempel, E., Einasto, M., Vennik, J., & Müller, V. 2008, *A&A*, 479, 927
- Toribio, M. C., Solanes, J. M., Giovanelli, R., Haynes, M. P., & Martin, A. M. 2010, *ApJ*, in press (Paper II)
- Verley, S., et al. 2007, *A&A*, 472, 121
- Wang, J., et al. 2010, *MNRAS*, in press (astro-ph/1011.0829)
- West, A. A., Garcia-Appadoo, D. A., Dalcanton, J. J., Disney, M. J., Rockosi, C. M., & Ivezić, Ž. 2009, *AJ*, 138, 796

Zwaan, M. A., Briggs, F. H., Sprayberry, D., & Sorar, E. 1997, ApJ, 490, 173

Table 1. Distributions of ALFALFA and SDSS Galaxies in Radial Velocity Bins

Δv_{hel} (km s ⁻¹)	ALFA N_{gal}	ALFA/SDSS-phot N_{gal} (%)	ALFA/SDSS-spec* N_{gal} (%)	SDSS-spec N_{gal}	SDSS-spec/ALFA N_{gal} (%)
< 500	71	65 (91%)	50 (77%)	37	18 (49%)
500–4000	1523	1417 (93%)	1024 (72%)	1702	910 (53%)
4000–7500	3104	3066 (99%)	2529 (82%)	5102	2420 (47%)
7500–11,000	3385	3353 (99%)	2947 (88%)	9196	2853 (31%)
11,000–14,500	2280	2261 (99%)	2025 (89%)	9341	1977 (21%)
14,500–18,000	876	871 (99%)	807 (93%)	10187	787 (8%)
Totals	11,239	11,033 (98%)	9382 (85%)	35,565	8965 (25%)

Note. — Numbers in different velocity bins of ALFALFA galaxies (ALFA), H I detections with a photometric counterpart in SDSS (ALFA/SDSS-phot), and H I detections with a spectroscopic counterpart in SDSS (ALFA/SDSS-spec*), regardless of the morphological classification of their SDSS spectrum, in the region where both catalogs overlap. SDSS-spec designates the sample of SDSS sources with both photometric and spectral morphology classified as of galaxy type and having $r < 17.77$ mag, while SDSS-spec/ALFA identifies the subset of SDSS-spec galaxies with an ALFALFA counterpart. Fractions are given in parentheses.

Table 2. H I Control Samples

Acronym	N_{gal}	Selection criterion
LDE	5647	Low local galaxy density environment ($\rho_6 \leq 0.5 h^3$ galaxies Mpc ⁻³)
IG1	461	Isolation inferred from spectro-photometric information (Equation 9)
IG2	236	Isolation inferred from photometric information (Equation 10)

Table 3. ALFALFA Detections and SDSS Galaxies in LDEs Obeying Mal09’s Criterion

Δv_{hel} (km s ⁻¹)	SDSS-spec		SDSS-spec/ALFA	
	N_{gal}	N_{disk} (%)	N_{gal}	N_{disk} (%)
2000–7000	2063	1893 (91.7%)	1330	1257 (94.5%)
7000–11,000	3912	3403 (87.0%)	1529	1411 (92.3%)
11,000–18,000	7889	6472 (82.0%)	1555	1414 (90.9%)
Totals	13864	11768 (84.9%)	4414	4082 (92.5%)

Note. — Numbers in different velocity bins of SDSS-spec and SDSS-spec/ALFA galaxies (the definition of these subsets can be found in the caption of Table 1) in low density environments satisfying Mal09’s disk-identification criterion (N_{disk}). We only consider galaxies with (r -band) Sérsic index $0.5 < n < 5.9$ and axis ratio $0.15 < b/a < 1$. Fractions are given in parentheses.

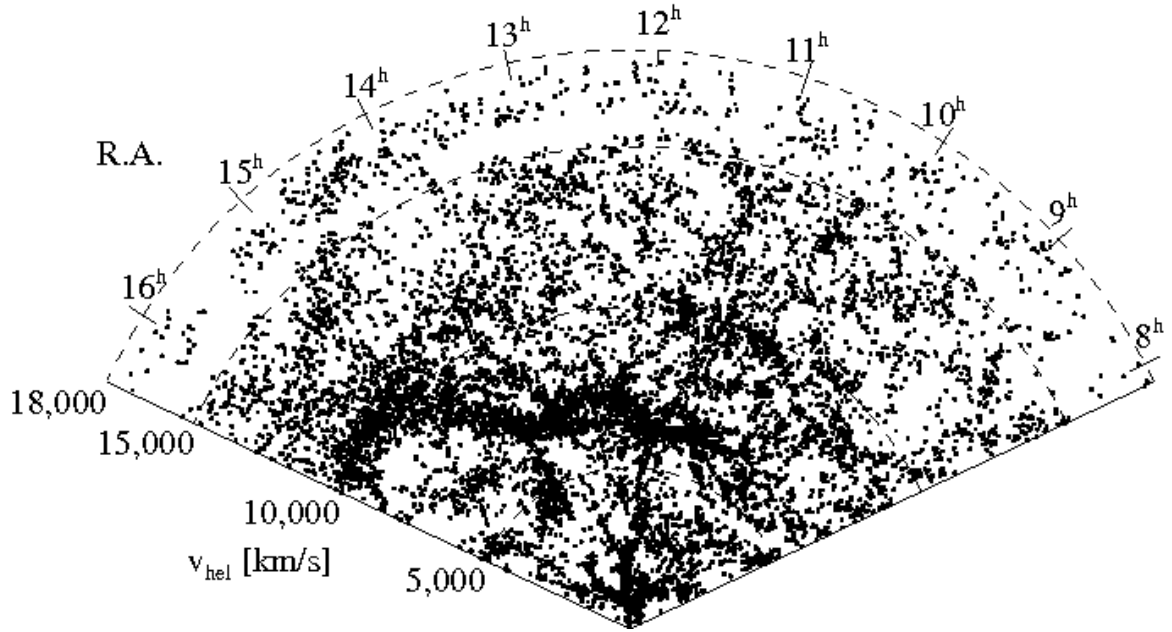


Fig. 1.— Sky distribution of ALFALFA sources in the sky region where it overlaps with the SDSS DR7 spectroscopic survey.

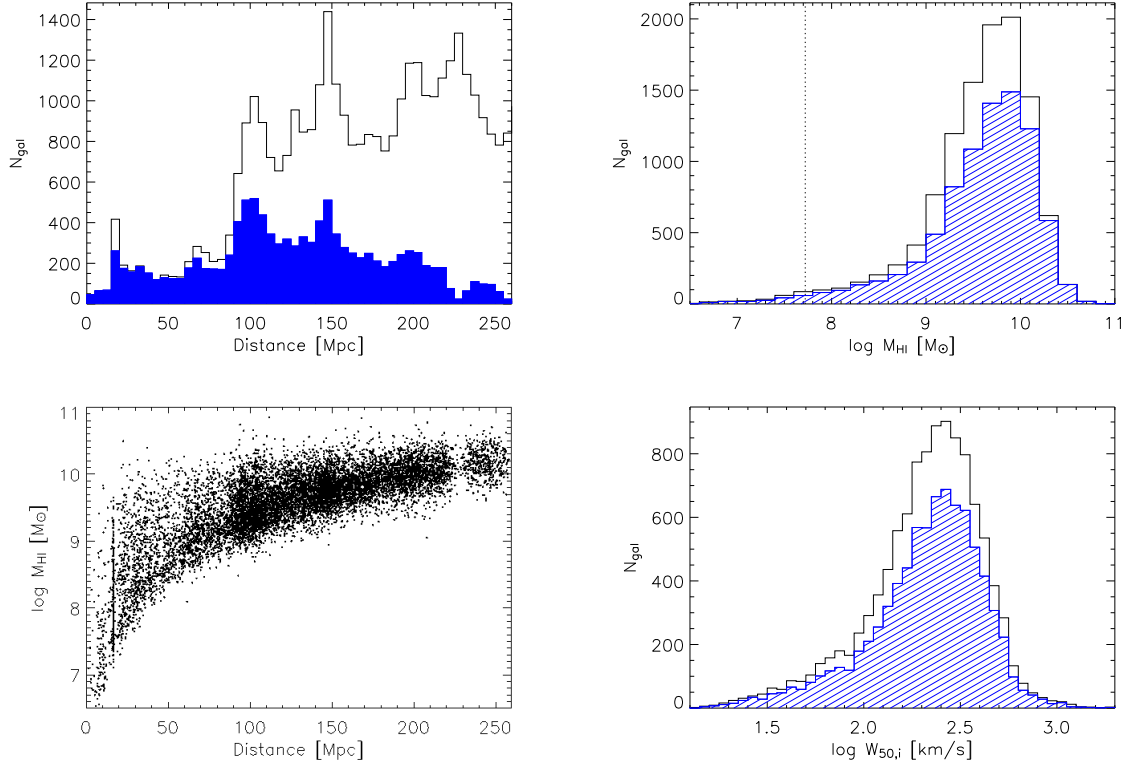


Fig. 2.— Top-left: Distribution of ALFALFA (solid-blue histogram) and SDSS (black line) galaxies according to CMB distance (see Section 2.2). Top-right: HI mass distribution for ALFALFA detections (solid line). The distribution for the ALFALFA code 1 detections is represented by the dashed-blue histogram. The vertical dotted line corresponds to the minimum HI mass detected at 2000 km s^{-1} . Bottom-left: Spänhauer diagram for the objects in the ALFALFA catalog. Bottom-right: Inclination-corrected velocity width distribution (black line) for ALFALFA sources. The dashed-blue histogram depicts the distribution arising from code 1 objects only.

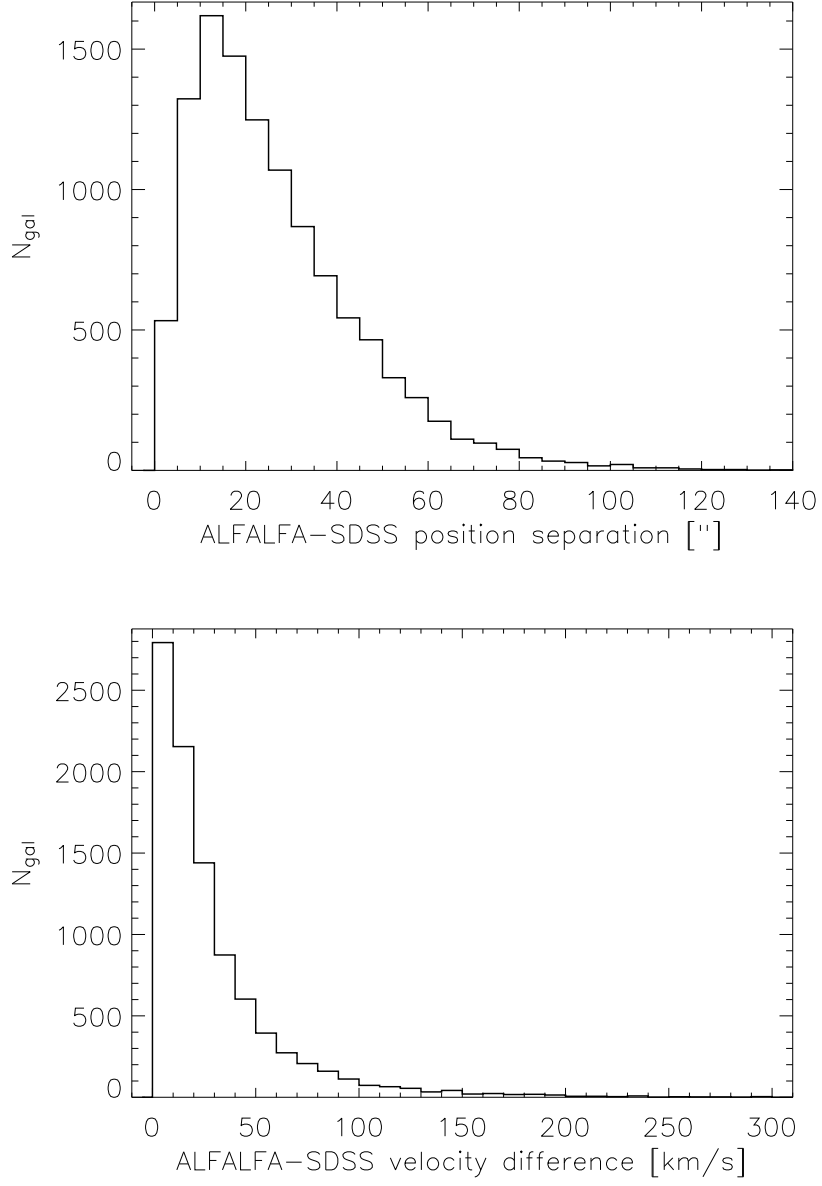


Fig. 3.— Top: Distribution of differences between HI and SDSS sky positions for ALFALFA detections with a SDSS photometric counterpart. Bottom: Distribution of differences between HI and SDSS radial heliocentric velocities for ALFALFA detections with a SDSS spectroscopic counterpart.

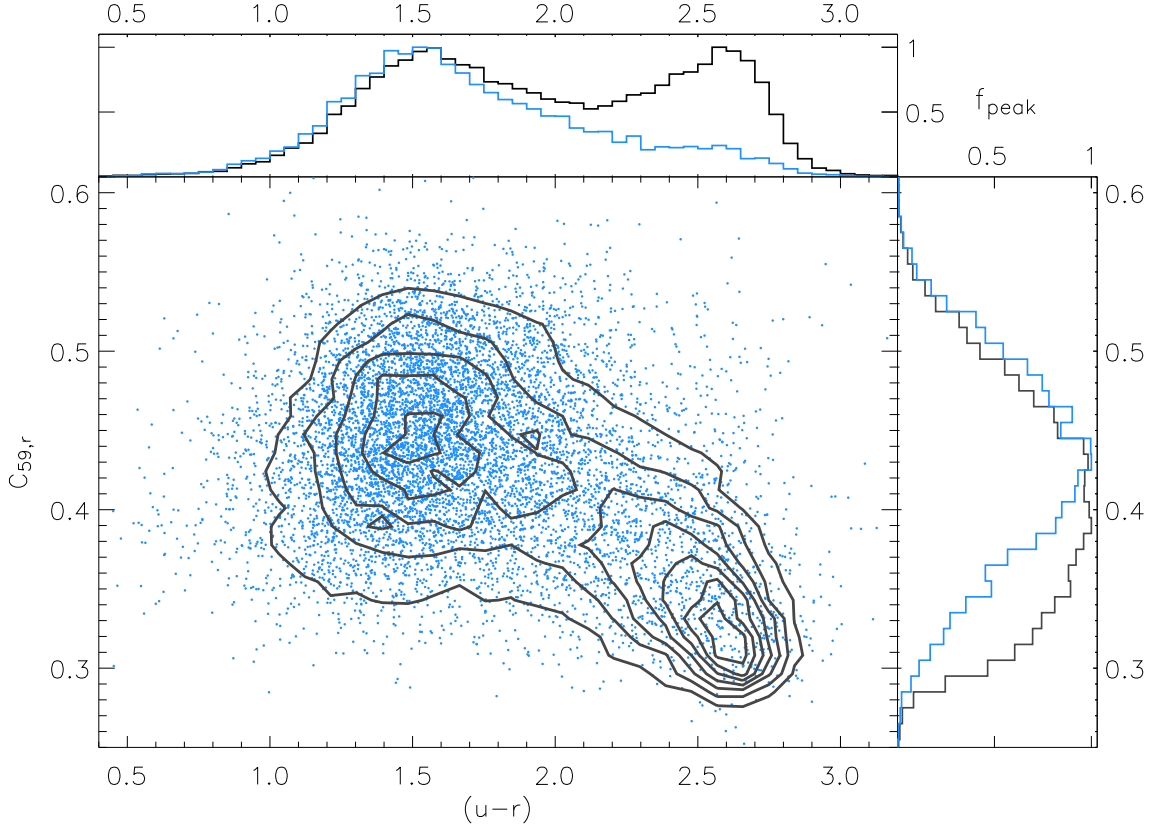


Fig. 4.— Inverse light concentration index in the r band, $C_{59,r}$, versus $(u-r)$ color for members of the SDSS spectroscopic sample (black contours), and for those that are also ALFALFA detections (blue dots). Contours are spaced at intervals of 40 density units. The histograms on the right and top sides of the main panel show the distributions of the corresponding variables for these two data sets normalized to peak values.

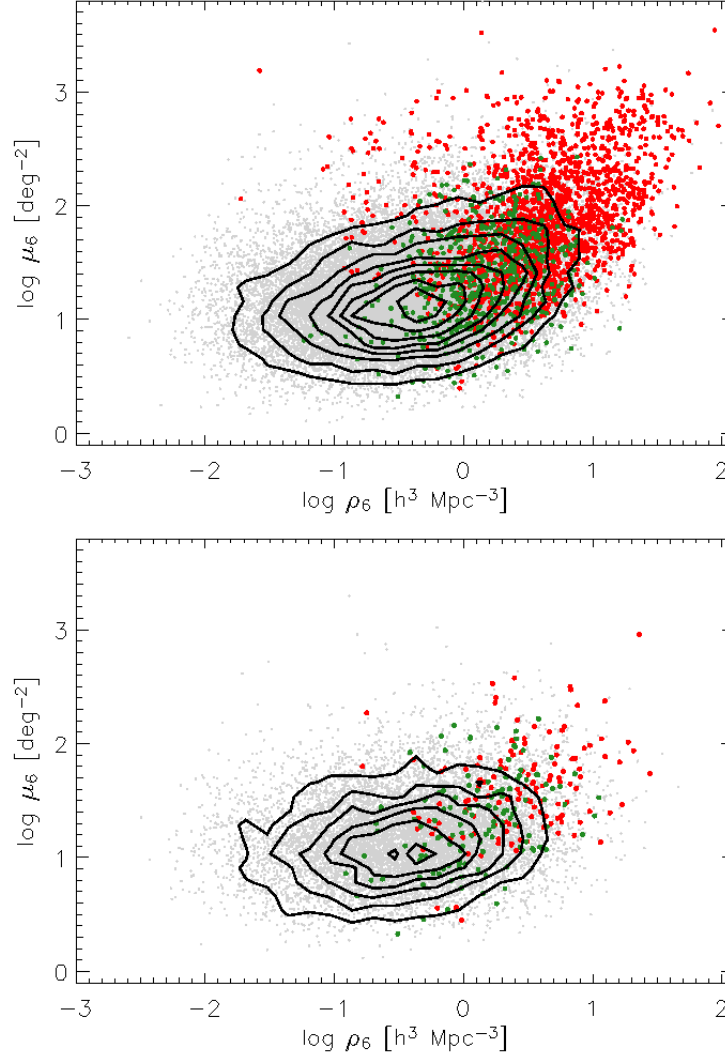


Fig. 5.— Three-dimensional local density of galaxies derived using the distance to the 6th-nearest neighbor, ρ_6 (Equation 5), in the SDSS-spec data set vs. the corresponding projected density estimator, μ_6 (Equation 7), calculated from the 2MASS galaxy distribution in the K_s band down to 13.0 mag. We show results for SDSS-spec galaxies (top) and for ALFALFA detections (bottom). Contours are spaced at intervals of 50 (top) and 25 (bottom) density units. Galaxies classified as cluster members are drawn using big red dots, whereas green dots are used for the clusters’ outskirts population. In these plots, the values of ρ_6 for cluster members have not been corrected for redshift-space distortions.

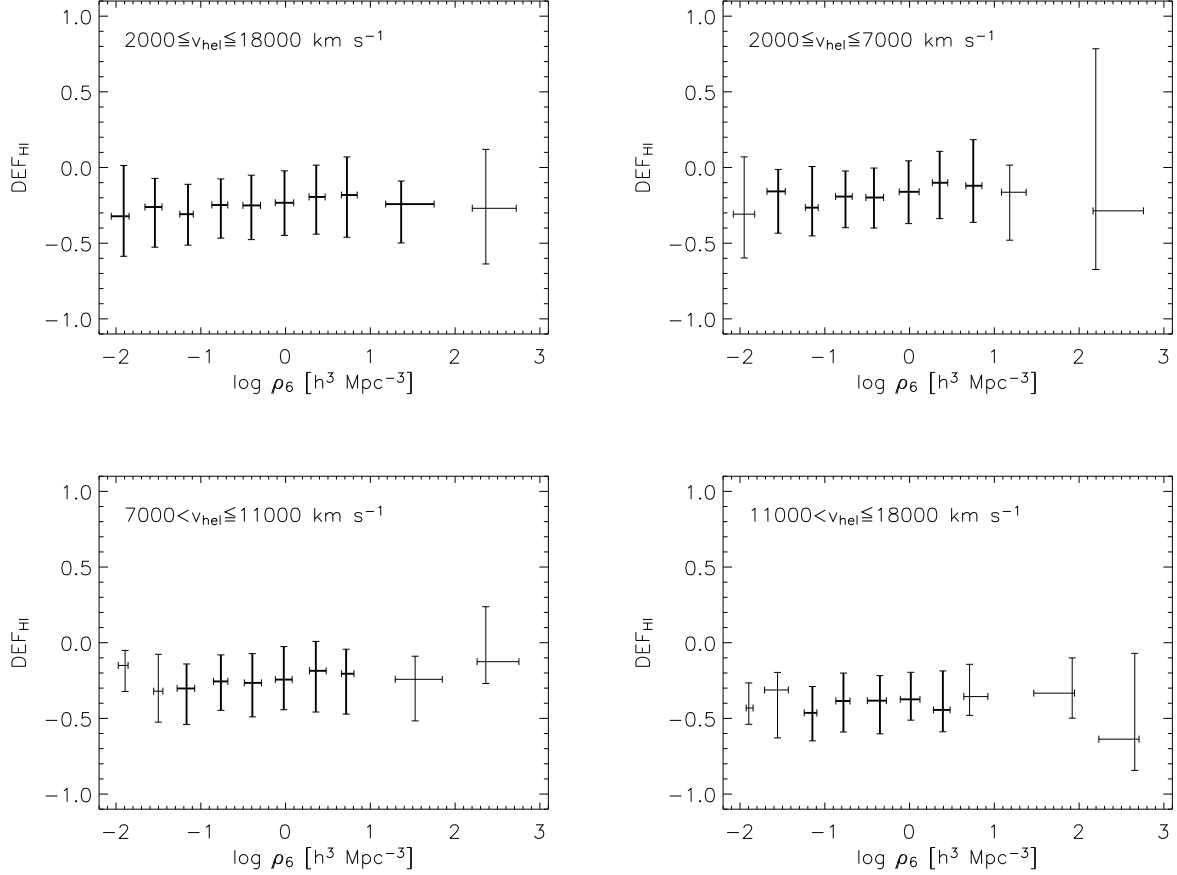


Fig. 6.— Medians and inter-quartile ranges of the HI-deficiency parameter, DEF_{HI} , calculated using the standards of HI content given by Solanes et al. (1996), in equal bins of $\log \rho_6$ up to $\log \rho_6 [h^3 \text{ Mpc}^{-3}] = 1$. Above the latter value, only two larger equal bins are considered. Results for bins containing less than 25 objects are represented by thinner lines. The different panels are for ALFALFA detections of Sa–Sd type with radial heliocentric velocities between 2000 and 18,000 km s^{-1} (top-left), as well as for the three subintervals: 2000–7000 km s^{-1} (top-right), 7000–11,000 km s^{-1} (bottom-left), and 11,000–18,000 km s^{-1} (bottom-right).

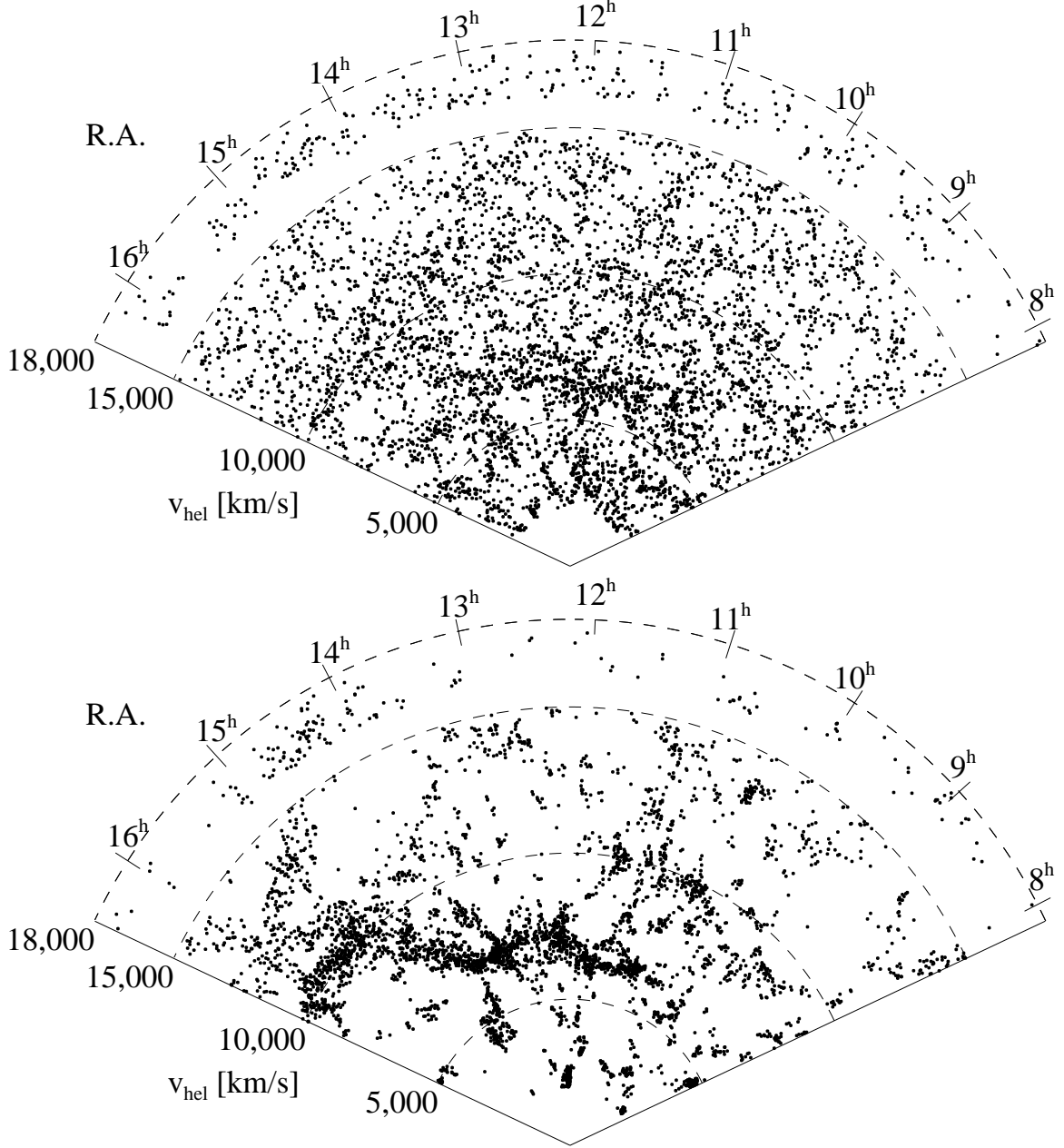


Fig. 7.— Wedge diagrams in right ascension of the LDE sample of ALFALFA galaxies defined in the text (top) and of those ALFALFA HI detections lying in environments of higher density in the same sky region (bottom).

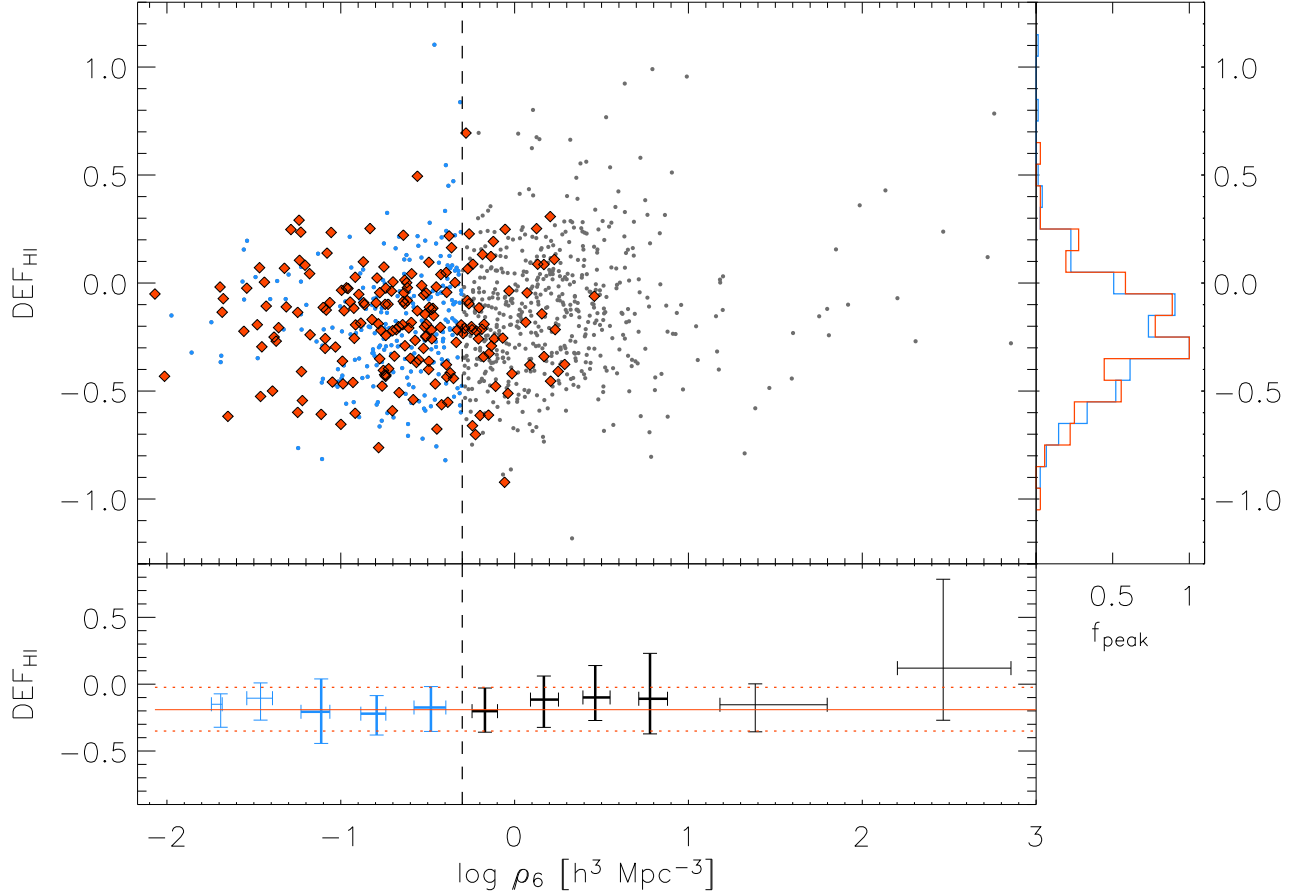


Fig. 8.— Top: DEF_{HI} , calculated using the standards of H I content given by Solanes et al. (1996), vs. $\log \rho_6$. ALFALFA detections of Sa–Sd type selected by a spectro-photometric isolation criterion (IG1 subset; see text) are shown as big red diamonds. Dots represent all Sa–Sd ALFALFA detections spanning the same range of g -band magnitudes than the isolated objects ($15.5 \leq g_i \leq 19.5$). The subset of galaxies to the left of the vertical dashed line (LDE galaxies) is highlighted in blue color. Bottom: Corresponding medians and inter-quartile ranges of DEF_{HI} . Results for all the ALFALFA galaxies depicted in the main panel are represented by error bars in equal bins of $\log \rho_6$ up to $\log \rho_6 [h^3 \text{Mpc}^{-3}] = 1$, while above the latter value only two larger equal intervals are considered. Results for bins containing less than 25 objects are drawn using thinner lines. The red horizontal lines show the median (solid) and upper and lower quartiles (dotted) of DEF_{HI} inferred from all the galaxies in the IG1 subset. Right: Histograms normalized to peak values showing the distribution of DEF_{HI} for the subsets of LDE galaxies (blue) and isolated sources (red). Similar results are obtained for the IG2 subset.

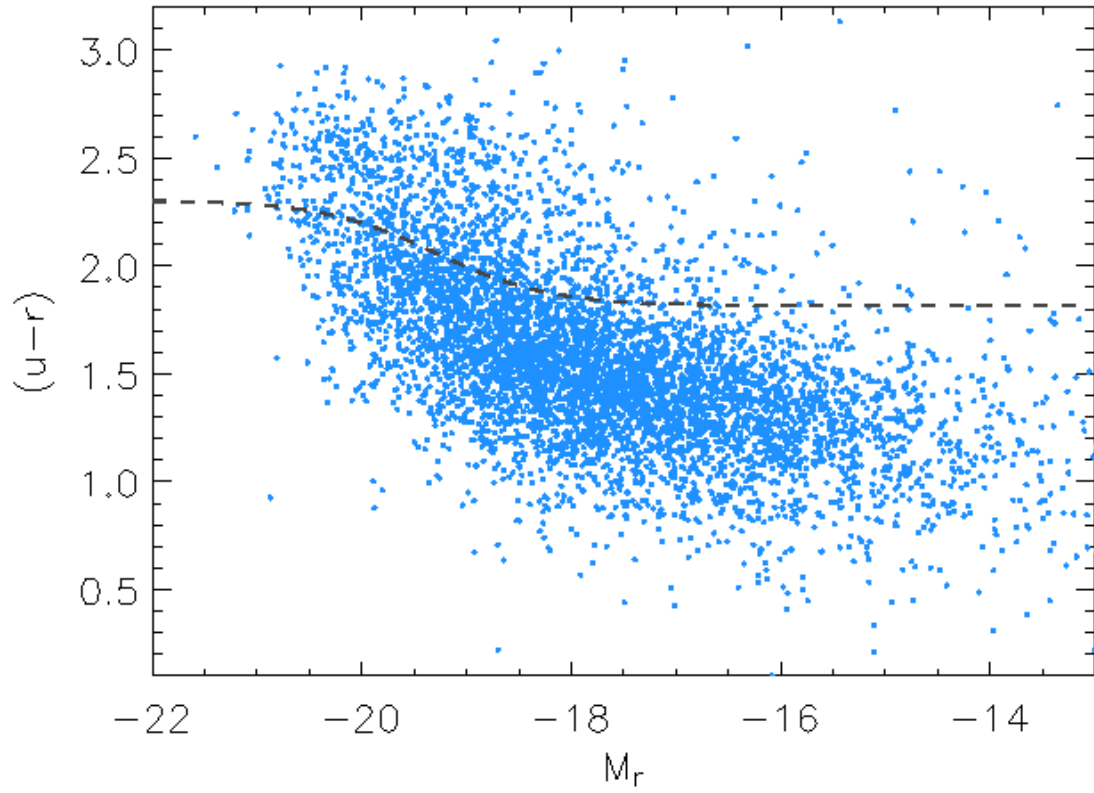


Fig. 9.— $(u-r)$ color versus absolute magnitude in the r -band for ALFALFA galaxies in the LDE subset. The dashed curve represents the red-blue population separator adopted by Baldry et al. (2004). Colors are uncorrected for internal extinction.

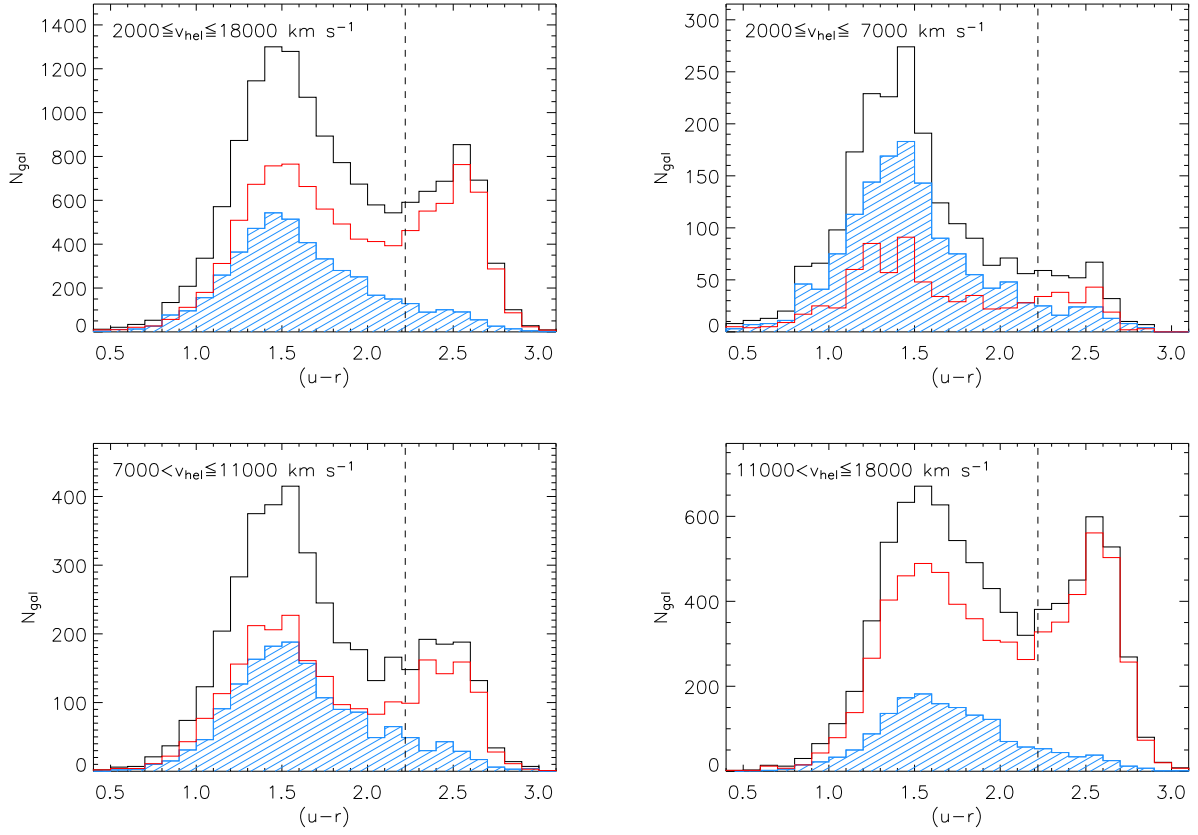


Fig. 10.— $(u-r)$ color distribution for SDSS galaxies in LDEs for different bins of radial heliocentric velocity. The global histograms are split into HI detections (dashed blue) and non-detections (red line). The vertical dashed line represents the red-blue population separator adopted by Strateva et al. (2001).

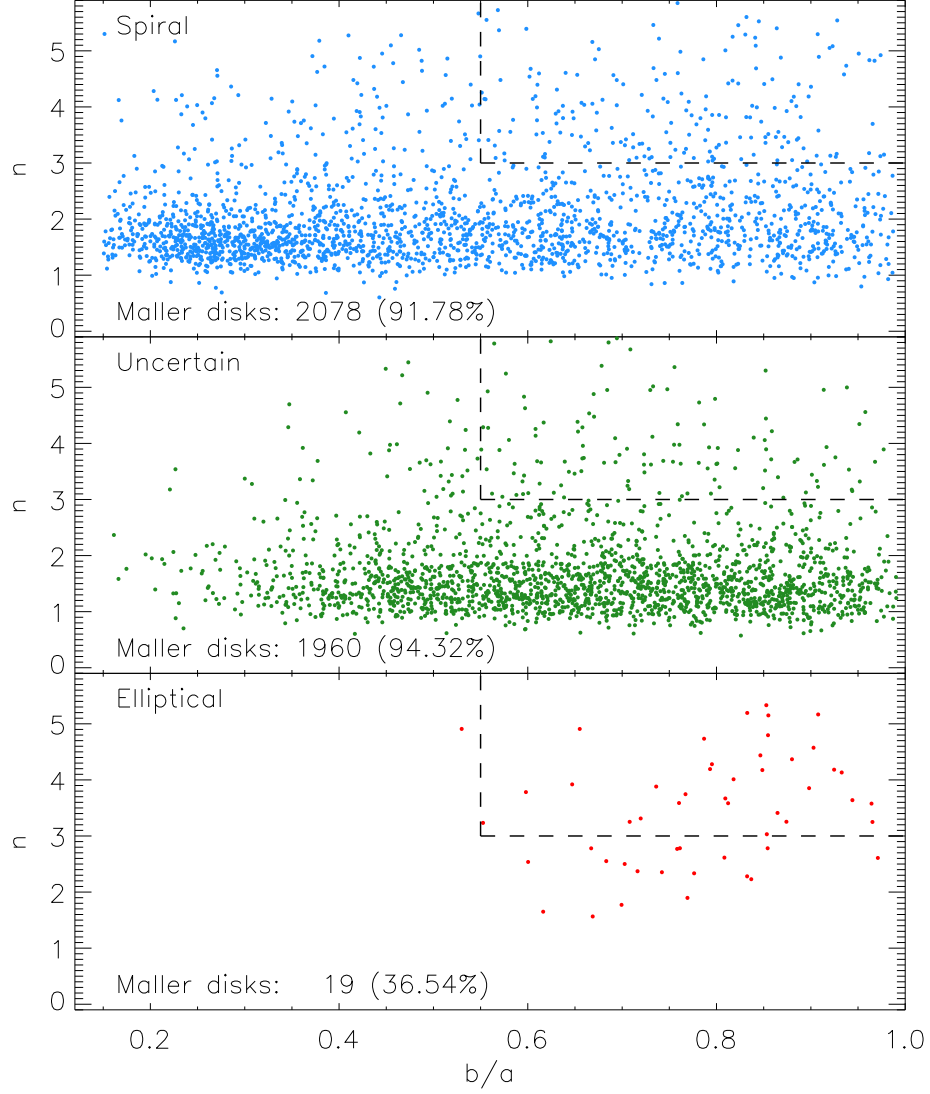


Fig. 11.— Distribution of GZ1 morphologies for ALFALFA LDE detections in the Sérsic index-observed axis ratio space (r -band): Spiral (top), Uncertain (middle) and Elliptical (bottom). In each panel, the top-right corner region delimited by dashed lines encompasses those objects not verifying Mal09’s disks criterion.

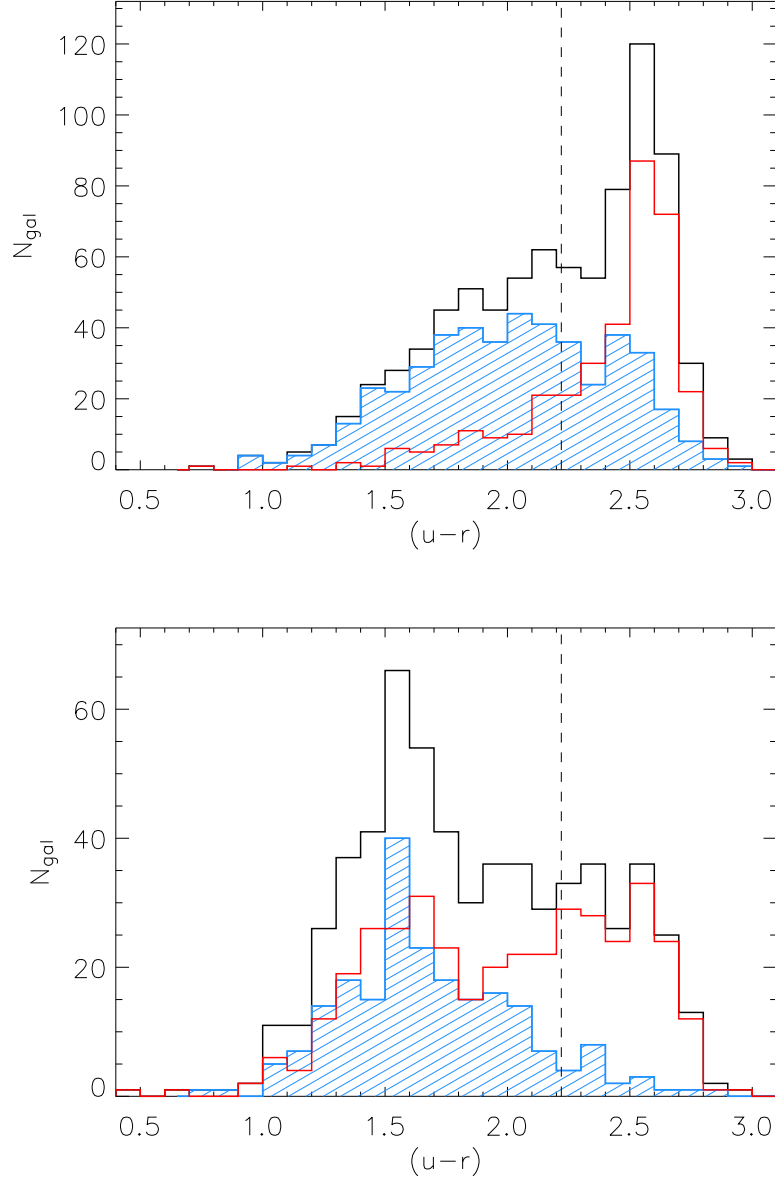


Fig. 12.— $(u - r)$ color distribution for SDSS galaxies obeying the spectro-photometric (top) and photometric (bottom) isolation criteria (see text). The global histograms are split into HI detections (dashed blue) and non-detections (red). The vertical dashed line represents the red-blue population separator adopted by Strateva et al. (2001).

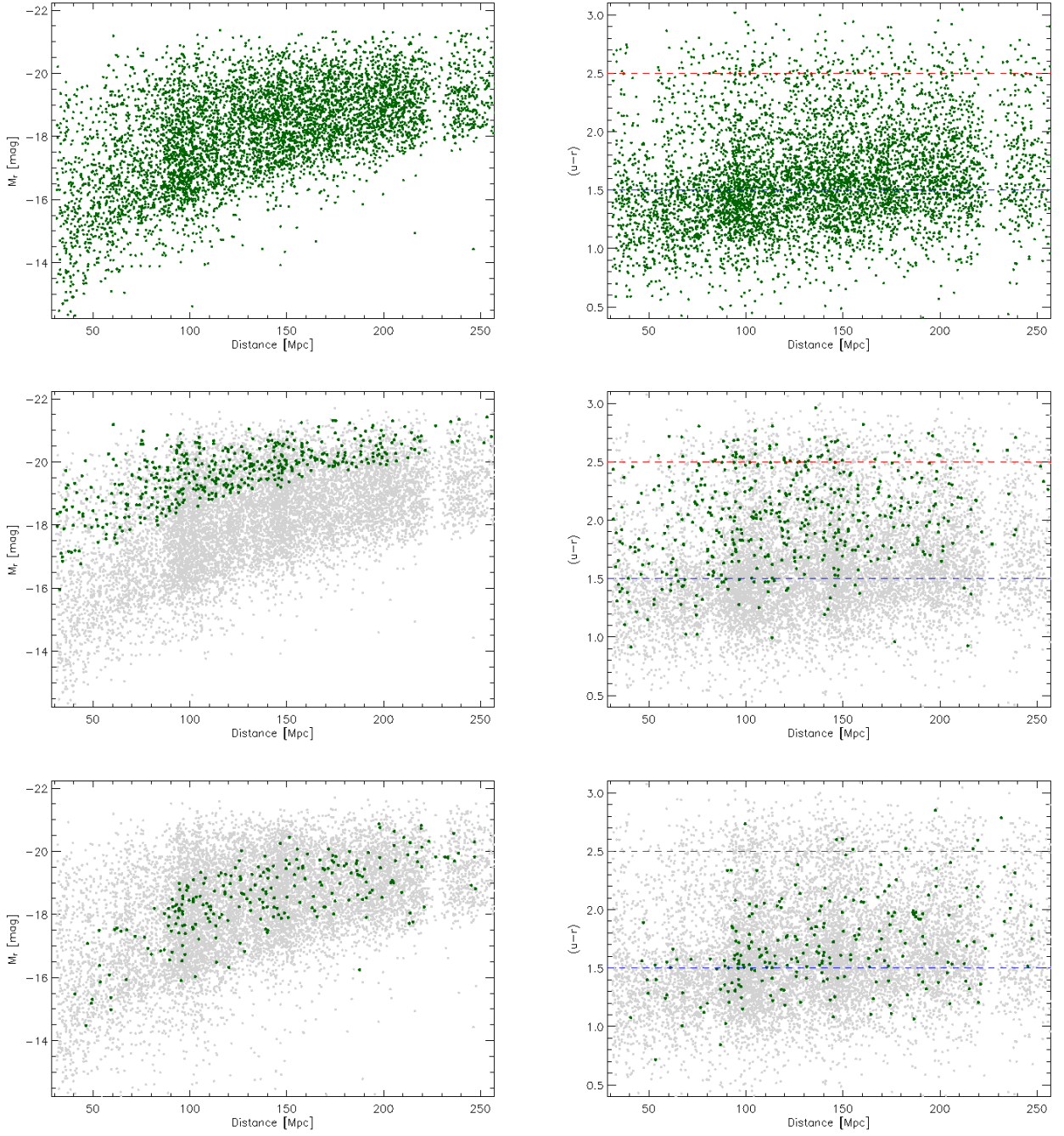


Fig. 13.— Absolute magnitude M_r (left) and $(u-r)$ color (right) vs. CMB distance for the three ALFALFA control samples: LDE (top), IG1 (middle), and IG2 (bottom). The member galaxies are drawn as green dots, whereas the middle and bottom plots also depict all the ALFALFA detections as gray dots. The horizontal dashed lines in the $(u-r)$ color plots show the peak values corresponding to the red and blue populations (see also Figures 10 and 12).